SIGNAL DETECTION THEORY IN THE STUDY OF NOCICEPTIVE AND PAIN PERCEPTION PROCESSES

CHEE-WEE TAN

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

QUEEN MARGARET UNIVERSITY
EDINBURGH

2008

Abstract

Signal detection theory (SDT) measures (discriminability and response bias) have been proposed to be valid for determining pain perception changes. The construct validity of SDT measures applied to pain perception studies has been questioned on three grounds: interpretation, methodology and theory.

Multiple interpretations are possible for the combinations of discriminability and response bias change when the magnitude-rating scale is used for pain perception studies. This is resolved by utilising the confidence-rating scale. The problem of comparability of results between the two scales is bridged by Irwin & Whitehead's (1991) common analytical framework. The results of this thesis supported the framework's prediction that both scales are comparable. Therefore, the confidence-rating scale was used for all studies within this thesis for interpretational clarity. Response bias data were not analysed in this thesis due to data artefacts created by correction methods for zero proportions in response categories.

Methodologically, the construct validity of discriminability is influenced by the research design and procedures. Therefore, the following procedures were adopted to address weaknesses in previous studies. The one-interval confidence-rating task was used with a six-category confidence-rating scale and post-trial feedback. Based on a methodological study conducted within this thesis, the trial number was pragmatically reduced from 40 trials to 17 trials per stimulus intensity. This trial number reduction would not alter the mean and variance of the data sufficiently to influence the outcome of inferential statistical testing performed. Due to the novel use of the Quantitative Sensory Testing machine for the signal detection study procedures, accuracy and precision study on the machine was performed. This thesis found that the accuracy, repeatability and reproducibility of the machine in generating noxious thermal stimuli is excellent for the purposes of this thesis. Machine error is eliminated as a major source of variance for the thesis results.

Theoretically, critics have challenged the construct validity of discriminability as an indicator of pain perception alteration. This thesis examined this issue in two separate contexts: 1) discriminability change as a correlate of local anaesthesia and, 2) discriminability as a correlate of psychological factors (depression and anxiety) in chronic low back pain (CLBP) sufferers. The results failed to establish the construct validity of discriminability for both contexts. However, the higher discriminability in CLBP sufferers compared to healthy individuals is in contrast to past research and warrant further investigation.

This thesis addressed the construct validity issues through theoretical, methodological and interpretational modifications. A more robust analysis of the construct validity issue was facilitated. Caution is recommended on the use of discriminability as a pain perception measure until the construct validity issue has been satisfactorily resolved.

Acknowledgements

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

Thanks go to:

Dr. Shea Palmer(supervisor)	*– for being such a brilliant, long-suffering, patient and positive supervisor.*
Dr. Denis Martin (supervisor)	*–for the uncanny gift of pointing out my grammatical mistakes in a multitude of ways (especially the funny ones) and all your encouraging words.*
Dr. Patricia Roche (supervisor)	*–for administering the last stages of my PhD.*
Mr. John Veto	*–for advising an EMLA® 'virgin' on the intricacies of the 'stuff'.*
Dr.Crawford Clark *Dr. Michael Hautus* *Prof. John Irwin* *Dr.Gary Rollman*	*– for their invaluable expert input.*
Western General Hospital Physiotherapy Department	*- for use of their premises in conducting my research and facilitating access to patients.*
Dr.Cathy Bulley *Dr.Fiona Coutts* *Dr.Frederike van Wijck*	*– for the encouragement, chats, free coffee, and laughs.*

^^^

Most of all:
Adeline, my wife
For the patience and love you've shown despite the trying times. My life belongs to you again.
'Place me like a seal over your heart, like a seal over your arm; for love is as strong as death…it burns like a blazing fire…many waters cannot quench love' (Song of Songs 8:6-7, NIV).

and Yulei, my first-born
For all the hugs and cuddles for Papa (with large amounts of protest, of course) and our one-worded debates on whether 'Yellow' is 'Green'. A philosopher in the making.

Mami & Papa
I finally did it!

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

Contents

CD content

Chapter 5 raw data: chapter_5_raw_data.xls (Microsoft Excel file)

Chapter 6 raw data: chapter_6_raw_data.sav (SPSS database file)

Chapter 7 raw data: chapter_7_raw_data.sav (SPSS database file)

Chapter 8 raw data: cgapter_8_raw_data.sav (SPSS database file)

Notations and abbreviations

α	Alpha level.
A'	An estimate of the area under the ROC based on a single point in ROC space.
A_g	An estimate of the area under the ROC based on more than one point in ROC space.
ANOVA	Analysis of variance.
A_z	The area under a linear ROC function.
β	The likelihood ration for two Gaussian distributions.
B	A response bias measure based on the geometry of ROC space.
B'_H	A response bias measure based on the geometry of ROC space.
B''	A response bias measure based on the geometry of ROC space.
BDI	Beck Depression Inventory.
BPD	Borderline personality disorder.
CLBP	Chronic low back pain.
CRT	Confidence-rating task.
d'	Discriminability measure assuming equal-variance distributions.
d_a	Discriminability measure assuming unequal-variance distributions and using the root-mean-square standard deviation.
dB	Decibel.
d_e	Discriminability measure assuming unequal-variance distributions and using the average standard deviation.
DSM	Diagnostic and Statistical Manual of Mental Disorders.
D_{YN}	The minimum distance between the origin to the linear ROC function.
c	A response bias measure in z-units from the equal-bias point.
°C	Degrees Celsius, a unit of measurement for temperature.
CDF	Cumulative Discriminability Function
CLBP	Chronic low back pain
e	The irrational number with a value of 2.718 (4 significant figures).
EMLA	Eutectic Mixture of Local Anaesthetics
F	False alarm.
fMRI	Functional Magnetic Resonance Imaging
FORCE	Function of Replications Combined Estimation.
H	Hit.
IASP	International Association for the Study of Pain.
INDSCAL	Individual Differences Scaling.
ISI	Interstimuli interval.
ISO	International Organization for Standardisation.
jnd	Just noticeable difference.
k	Weber's constant.
K	Kelvin, a unit of measurement for absolute temperature.
ln	Natural logarithm.
Mdn	Median.
MRT	Magnitude-rating task.
μ_i	Mean of the normal probability density of stimulus class i.
NIST	National Institute of Standards and Technology.
Φ	The probability density function of the normal distribution.
$P(A)$	An estimate of the area under the ROC based on more than one point in ROC space.

$P(\cdot\|\cdot)$	Conditional probability.
QST	Quantitative Sensory Testing
ROC	Receiver Operating Characteristic
σ_k	Standard deviation of the variable k.
s	Slope of the linear ROC function.
SDT	Signal detection Theory.
S.E.	Standard error.
STAI	State-Trait Anxiety Inventory.
TENS	Trancutaneous Electrical Nerve Stimulation.
$z(F)$	False alarm rate in z-units.
$z(H)$	Hit rate in z-units.
$z(p)$	Probability transformed to z-units.

Referencing style:
All citations and references within this thesis follow the style of the Publication Manual of the American Psychological Association, 5th edition (American Psychological Association, 2001).

List of figures

List of tables

Chapter 1
Introduction

1.1 Introduction

Knowledge about pain perception has advanced within the last half century. This is in part due to the assimilation of methodologies from other disciplines of study to complement the unique investigational paradigms used within pain perception research. The use of psychophysical methodologies is one such example.

1.2 Psychophysics in pain perception

Psychophysics is “the scientific study of the relation between stimulus and sensation” (Gescheider, 1997, p. ix). Although this definition is accurate in that the construct of stimulus and sensation has been mentioned, psychophysical research goes beyond sensation to also encompass the study of perception. This is exemplified by pain perception researchers using psychophysical methods to elucidate nociceptive and pain perception mechanisms and processes. For example, Hardy, Wolff & Goodell (1947, 1948) used radiant heat as the source of noxious stimulus and increased the physical intensity of the stimulus in order to obtain the first noticeable change in stimulus intensity by the participant. This procedure was repeated for a range of stimulus intensities. This first noticeable change in stimulus intensity is termed the ‘just noticeable difference’ (Gecheider, 1997, p.398). Based on the number of just noticeable differences observed for a range of physical stimulus intensities, Hardy, Wolff & Goodell (1947, 1948) constructed a scale of noxious increments called the dol scale. Although this scale has now been superseded by other modern psychophysical methods, the application of psychophysics for the study of pain perception had been established. To further illustrate the influence of psychophysics on concepts within pain perception research, the construct of pain threshold is based partly on “the least experience of pain which a subject can recognize” or the level at which “50% of the stimuli would be recognized as painful” (International Association for the Study of Pain Task Force on Taxonomy, 1994). This latter definition is synonymous to the classical psychophysical concept of a threshold.

1.3 Psychophysical methods in pain research

There are numerous types of psychophysical methods and scales that may be used to study the relationship between the noxious stimulus and the associated response generated. Some examples of psychophysical methods used in pain perception research are the modified method of limits (Békésy, 1947), magnitude estimation (Stevens & Stevens, 1975), categorical scaling, cross modality scaling, master scaling (Berglund & Harju, 2003), category ratio scaling (Borg, 1998) and signal detection theory (Green & Swets, 1966).

Clark (1974, 1994) proposed that signal detection theory (SDT) was a useful method for researching pain perception. One of the advantages of SDT in the context of pain perception research is that it provides two indices, the 'discriminability' and the 'response bias'. These indices correspond respectively to the sensory or perceptual separability of the perceived stimuli and the decisional criteria by which participants make their responses. The separation of the participant's capacity to discriminate between stimuli and the decisional criteria is noted to be the strength of SDT over some other psychophysical methods for researching sensation and perception (Green & Swets, 1966; Swets, 1996, Macmillan & Creelman, 2005). For the non-SDT psychophysical methods, the decisional criteria are usually confounded with the index generated for determining the participant's sensory or perceptual ability. Since the magnitudes of the index represent the sensory or perceptual ability of the participants, this may lead researchers to conclude wrongly that the participants possess different sensory or perceptual abilities.

Due to analytical artefacts created during the data reduction procedures for response bias, this thesis excluded the analysis of response bias results to prevent erroneous conclusion drawn from the results. The artefacts are generated in part due to the presence of non-utilisation of some categories with the rating scale used. Chapter 4 outlines how artefacts may be produced for response bias and the implications on result interpretation.

A second strength of SDT lies in the adaptability of its methodology to include data from other psychophysical methods for analysis. The advantage of this broad data

analytical approach allows for the approximate comparison of data from different psychophysical methods. For example, with some modifications to the data reduction strategy for response data obtained through magnitude estimation, the reduced data may be compared with the response data obtained from SDT methodology (Cohen & Lecci, 2001). This particular utilisation of SDT goes beyond its use as a methodological tool for elucidating the sensory or perceptual ability of participants. It also highlights the potential of SDT as a unifying framework for analysing response data from different methodologies. Chapter 2 describes the theory and method underlying the use of SDT and Chapter 5 outlines a unifying framework by Irwin & Whitehead (1991) for comparing response data between different methods.

Researchers have used SDT for investigating the antinociceptive properties of pharmacological products (Janal, Colt, Clark & Glusman, 1984), transcutaneous electrical nerve stimulation (McCreery & Bloedel, 1978), the influence of affect on pain perception (Dworkin, Clark & Lipsitz, 1995), and the pain perception ability of chronic pain sufferers (Yang, Richlin, Brand, Wagner & Clark, 1985). Chapter 3 reviews some of the SDT pain perception literature relevant to the research question and design for this thesis.

Signal detection theory is derived from statistical decision theory, the general foundational theory for most inferential statistical hypothesis testing (Egan, 1975). Signal detection theory was specifically named because of its application in modeling the detection behaviour of signalmen manning radar equipment (Macmillan & Creelman, 2005, p.22-24; Tanner & Swets, 1954). Detection is the ability to filter out true signals from the constant background noise that is present in the information presented to the person making the decision, in this example, the decision maker is the signalman (Green & Swets, 1966, Swets & Pickett, 1982, Macmillan & Creelman, 2005). However, the application of SDT was later extended to psychological research that investigated the participant's ability to differentiate between two physical stimuli of different intensities. This process of differentiating between two physical stimulus intensities is termed discrimination. For example, participants may be asked to discriminate between two or more auditory signals of varying intensities (Braida & Durlach, 1972). The similarity between the two processes of detection and discrimination is the involvement of a participant receiving information from the

external environment. This information is then processed internally and a decision is made regarding the nature of the information received. As decision making is an important facet of signal detection theory, researchers who have adopted signal detection theory for investigating pain perception have named it ‘Sensory Decision Theory’ (Chapman, Chen, & Bonica, 1977; Clark, 1969). However, in this thesis, the term ‘Signal Detection Theory’ will be adopted for consistency. The reason for this preferred term is based on convention in the general psychophysical literature which has used this term and related variants, including ‘Detection Theory’ and ‘Theory of Signal Detection’. The use of ‘Signal Detection Theory’ also differentiates it from ‘Statistical Decision Theory’ in relation to the area of application, the latter being associated with statistical hypothesis testing.

SDT is an established theoretical framework within sensory and perception research, for example in recognition memory (Rotello, Macmillan & Reeder, 2004) and vision (Eckstein, Thomas, Palmer & Shimozaki, 2000). However, its inception within pain perception research was more difficult. The crucial issue underlying this debate is the construct validity of SDT indices when used for investigating pain perception processes. One of the main debates surrounds the ambiguity in interpretation of the SDT indices for the study of pain perception processes. Opponents argued that the indices conveyed little information specific to pain perception (Rollman, 1977). This implied that the knowledge obtained about pain perception through SDT-related methodology could be potentially misleading. However, it was the soundness of the interpretation and not the theory itself that was questioned (Rollman, 1977). The debate can be divided into three slightly overlapping domains: Theory, methodology and definition/interpretation. Chapter 4 outlines and discusses the debate regarding the use of SDT for investigating pain perception.

Theoretical and methodological developments within SDT have proceeded in other sensory and perceptual disciplines. These developments have had little influence on the further evolution of SDT within pain perception research. More recently, Irwin & Whitehead (1991) proposed a framework that incorporated SDT with psychophysical theories by Laming (1984) and Braida & Durlach (1972). This framework was then applied to the study of a topical local anaesthetic (Irwin, Hautus, Dawson, Welch & Bayly, 1994). Chapter 6 outlines the framework proposed by Irwin & Whitehead

(1991). As mentioned previously, this SDT-based framework acts as a unifying approach for analysing response data from different methods. This thesis also examined the comparability of response data from two methods: the magnitude-rating method which has been used by most SDT studies in pain perception, and the confidence-rating method which is used in the studies within this thesis. In conjunction with the theoretical refinement offered by Irwin & Whitehead's (1991) framework, more understanding about the influence of certain analytical procedures within SDT has emerged. In particular, the issues of the number of stimulus trials administered to participants, the number of response categories used by participants for judgment and their contribution to the accuracy of the SDT indices is being understood (Hautus, 1995; Miller, 1996). Chapter 4 outlines some of these methodological issues and their influence on the methodology used in this thesis. Experiment D in Chapter 5 (Sections 5.20-5.25) describes a methodological study conducted for this thesis with the aim of comparing the amount of variance when the number of stimulus trials administered to participants is reduced from 40 to 17 trials.

It is important to examine the issue of construct validity of discriminability as a correlate of pain perception processes. The utility and usefulness of SDT indices for pain perception research requires the resolution of the construct validity issue. This thesis chose two specific research contexts for the examination of the construct validity issue of discriminability as a correlate in pain perception processes. The first context (Chapter 7) involved the induction of cutaneous local anaesthesia with the evaluation of discriminability change as an indication of analgesia. This research context examined the construct validity of discriminability under a known, localised reduction of cutaneous sensation. The second context (Chapter 8) involved the description and comparison of discriminability to noxious thermal stimuli, administered using a contact thermode, in CLBP sufferers compared to healthy individuals. This research context compared findings from previous SDT studies that have investigated discriminability in CLBP sufferers. This thesis also examined the correlation between the affective factors of depression and anxiety severity with discriminability within the second research context. The correlational analysis was performed to find out the amount of variance depression and anxiety severity contributes to discriminability. This gathered evidence for the examination of the

construct validity of discriminability as a correlate of psychological factors in pain perception processes for a clinical population.

1.4 Aims of thesis

This thesis has one general aim and four specific objectives:

General aim:

To gather evidence regarding the construct validity of discriminability as a correlate of pain perception processes.

The three domains of construct validity: theory, methodology and definition/interpretation, as commented on by critics of SDT applied to the study of pain perception, will be examined using appropriate evidence-gathering strategies within this thesis. The evidence-gathering strategies used in this thesis included literature reviews and the conduct of empirical studies.

Specific objective 1:

To incorporate some of the theoretical and methodological advances within SDT to the investigation of pain perception.

This objective examines the methodological domain of criticisms about the construct validity of discriminability as a correlate of pain perception processes.

Specific objective 2:

To use the analytical framework proposed by Irwin & Whitehead (1991) for investigating the comparability of the magnitude-rating task and the confidence-rating task.

This objective provides support to the interpretation clarity of discriminability as a correlate of pain perception processes. This aim thereby gathers evidence relating to the methodological and theoretical domains of criticisms about the construct validity of discriminability as a correlate of pain perception processes.

Specific objective 3:
To examine the construct validity of discriminability as a correlate of analgesia induced using topical local anaesthetic.

This objective gathers evidence regarding the theoretical domain of criticisms about the construct validity of discriminability as a correlate of pain perception processes.

Specific objective 4:
To examine the construct validity of discriminability as a correlate of pain perception processes associated with psychological factors (depression and anxiety) for CLBP sufferers.

This allows comparison between this thesis' results with results from previous SDT studies investigating pain perception processes for the above patient group. This aim gathers evidence regarding the theoretical domain of criticisms about the construct validity of discriminability as a correlate of pain perception processes.

Essentially, specific objectives 3 and 4 seek to accumulate evidence to either support or refute the proposed interpretation of the discriminability within two specific contexts of pain perception research: analgesia induced by topical local anaesthetic and psychological factors in pain perception.

The original contribution of this thesis is the incorporation of theoretical and methodological advances within a SDT framework to the study of pain perception. This is further supported by the verification of Irwin & Whitehead's (1991) proposed analytical framework within two different research contexts.

1.5 Structure of thesis

This thesis is structured as follows:
Chapter 2: This chapter describes the basic theory and concepts underpinning SDT. The role of the rating task, a psychophysical procedure used within this thesis' method, is also explained together with the associated SDT discriminability indices of d', d_a, A_z and $P(A)$.

Chapter 3: This chapter reviews some of the pain perception research using SDT methodology relevant to this thesis.
Chapter 4: This chapter explains and discusses some of the main criticisms and challenges regarding the interpretation of SDT indices within pain perception research. Several procedural issues for the conduct and analysis of data obtained are also discussed. The criticisms are categorised into three domains for discussion: theory, methodology and definition/interpretation.
Chapter 5: This chapter describes a calibration and precision study of the equipment (Somedic® Thermotest), used throughout this thesis, for the administration of the contact noxious thermal stimuli. It also describes a methodological study comparing the relative efficiency between two stimulus trial numbers.
Chapter 6: This chapter examines the comparability between participant responses obtained using the magnitude-rating task and the confidence-rating task. The SDT analytical framework proposed by Irwin & Whitehead (1991) is used for this examination.
Chapter 7: This chapter describes a study investigating the construct validity of discriminability as a correlate of analgesia. In this study, analgesia was induced by a topical local anaesthetic (EMLA®).
Chapter 8: This chapter describes a pseudo-experimental study investigating the construct validity of discriminability as a correlate of psychological factors in pain perception processes. The sample groups of chronic low back pain sufferers and healthy individuals were recruited.
Chapter 9: This chapter provides a general discussion of the findings for this thesis.

The intention of this structure is to provide a comprehensive overview of the research programme.

Chapter 2
Signal Detection Theory: Theory and methods

2.1 Overview of chapter

The techniques used by scientists to investigate sensation and perception are wide-ranging. Although psychophysical methods and procedures are fairly simple to administer, these techniques assume theoretical principles and models that are used to describe the behavioural data collected from the experiment. This also applies to signal detection theory (SDT) and its models.

The application of SDT to research problems requires an appreciation of the theoretical underpinnings of the various models that have been developed and their associated methods. This chapter will provide an introduction to the basics of SDT. The methods, discriminability measures and response bias measures of the 'yes-no' experiment and 'rating' experiment will be elaborated upon. This review will also outline the rationale for the specific procedures selected and applied in this thesis.

2.2 The fundamental decision problem

In the experimental or natural environment that the participant is placed into, each occurrence of an experiential state, which involves the combination of stimulus and response, is called an event (Swets, Tanner & Birdsall, 1961; Egan, 1975). For example, a pedestrian waiting to cross a road will experience visually the occurrence of one of two stimuli (i.e. the stimulus of a red traffic signal or the stimulus of a green traffic signal) before appropriate action is taken (i.e cross the road or not cross the road). In the most fundamental decision problem, where one of two stimuli occurred, the task of the participant is to state which stimulus occurred coupled with the appropriate response. As stated previously, the combination of stimulus and response is called an event.

A hypothetical example will be described below from which further elaboration of detection theory will draw.

In a detection experiment, a participant is required to detect if a signal has been presented from a background of noise. At timed intervals, two conditions can be presented to the observer: the first condition consists of only background noise with no signal presented, and the second condition consists of the noise plus the signal. When indicated, the participant will respond either 'No' (the signal has not been presented) or 'Yes' (the signal has been presented).

The experiment therefore concerns two types of stimuli and two types of responses. Accordingly, any one of four events may occur on each particular trial. When the participant correctly indicates a 'Yes' response when the signal is presented, the event is termed a 'hit'. Indicating a 'Yes' when the signal has not been presented, the event is termed a 'false alarm'. Responding 'No' when a signal is presented, the event is termed a 'miss'. And the last possible event of responding 'No' when no signal has been presented, the event is termed a 'correct rejection'. The possible events are summarised in a response matrix (Table 2.1) below.

	Response	
Stimulus	***"Yes"***	***"No"***
Signal	Hits	Misses
Noise	False alarms	Correct rejections

Table 2.1 Response matrix for the yes-no experiment. Four events are possible: Hits, misses, false alarms and correct rejections.

Assuming that 25 noise and 25 signal trials have been presented to the participant for detection. At the end of the experiment, a response matrix may be represented like Table 2.2.

	Response		
Stimulus	***"Yes"***	***"No"***	***Total***
Signal (S)	Hits (20)	Misses (5)	25
Noise (N)	False alarms (10)	Correct rejections (15)	25

Table 2.2 Response matrix for a yes-no experiment. The frequencies of responses to the particular stimulus are included within the matrix.

Although four events are possible for the response matrix, only two numbers are necessary to summarise the participant's proficiency in differentiating between the two stimuli. Conventionally, the 'hit' rate and the 'false alarm' rate are chosen to calculate the associated outcome measures within signal detection theory. The hit and false alarm rates are calculated as follows (Wickens, 2002, p.8):

$$Hit\ rate(H) = \frac{Number\ of\ hits}{Number\ of\ signal\ trials} \tag{2.1}$$

$$False\ alarm\ rate(F) = \frac{Number\ of\ false\ alarms}{Number\ of\ noise\ trials} \tag{2.2}.$$

For example, using data from Table 2.2, the Hit rate and the False alarm rate are calculated as follows:

$$H = \frac{20}{25} = 0.8$$

$$F = \frac{10}{25} = 0.4$$

The entire response matrix of Table 2.2 can be rewritten in terms of the response rates rather than their frequencies:

	Response		
Stimulus	***"yes"***	***"no"***	***Total***
Signal (S)	0.8	0.2	1.0
Noise (N)	0.4	0.6	1.0

Table 2.3. Response matrix for the yes-no experiment expressed in terms of proportions.

In sensory and pain perception studies, the research design usually involves two clearly perceptible signal stimuli of different intensities instead of one signal stimulus and one stimulus consisting of only background noise. Study designs adopting the two signal stimuli design will be termed 'discrimination experiments' in this thesis. When the events are interpreted in the context of the discrimination experiment, the hit rate indicates the probability estimate of the participant correctly identifying the higher intensity stimulus when it is presented, and the false alarm indicates the probability estimate of the participant incorrectly identifying the lower intensity stimulus as the higher intensity stimulus.

2.3 The discriminability index, *d'*

The objective of the experimenter is to find the participant's proficiency in detecting the signal from the background noise (or discriminating between two signal stimuli). The measure of proficiency is called the discriminability. In the case of the discrimination experiment, the participant's proficiency will be termed the discriminability in this thesis.

When a participant has perfect discriminability, the hit rate will be 1 and the false alarm rate will be 0. A completely insensitive participant will be unable to distinguish between the background noise and the signal stimuli. In this instance, the participant may obtain correct responses purely by chance. Therefore, the participant is said to be operating at chance level and so the hit and false alarm rates will be approximately similar. Practically, most situations will fall between the two extremes of perfect discriminability and complete insensitivity.

The above detection experiment example consisted of two stimuli, signal and noise. To decide which one of these stimuli has occurred, the participant is given some evidence or information in the form of the stimuli. The task of the participant is to make a decision if the evidence presented favoured one of the two hypotheses that the signal or the noise stimuli occurred. A graph may be drawn to represent the presentation probability of this evidence (Figure 2.1). Since Figure 2.1 characterises the internal representation of the problem or task facing the participant, the figure is also called the 'decision space' (Macmillan & Creelman, 2005, p.16). The x-axis of Figure 2.1 is the evidence variable. The form that the evidence variable takes is dependent on the context of the experiment. In the situation where the participant is required to distinguish between the temperature magnitudes of two thermal stimuli and determine which stimuli is more intense, the evidence variable would be the perceived magnitude of the temperatures. SDT assumes that the evidence or information as perceived by the participant is affected by random variation. This means that the stimuli may be described as a range of magnitudes as perceived by the participant. Therefore, this information as perceived by the participant may be represented by a Gaussian or normal distribution. The y-axis of Figure 2.1 represents the probability of the evidence occurring when either one of the two stimuli is presented. When the signal stimulus is presented, the participant may perceive the magnitude of the stimulus along any point under the Gaussian distribution of the signal stimulus (right-most distribution). The estimated probability that the signal stimulus is detected is represented on the distribution at the corresponding y-axis value. This description of the internal representation of the stimulus may also be applied to the noise stimulus (left-most distribution).

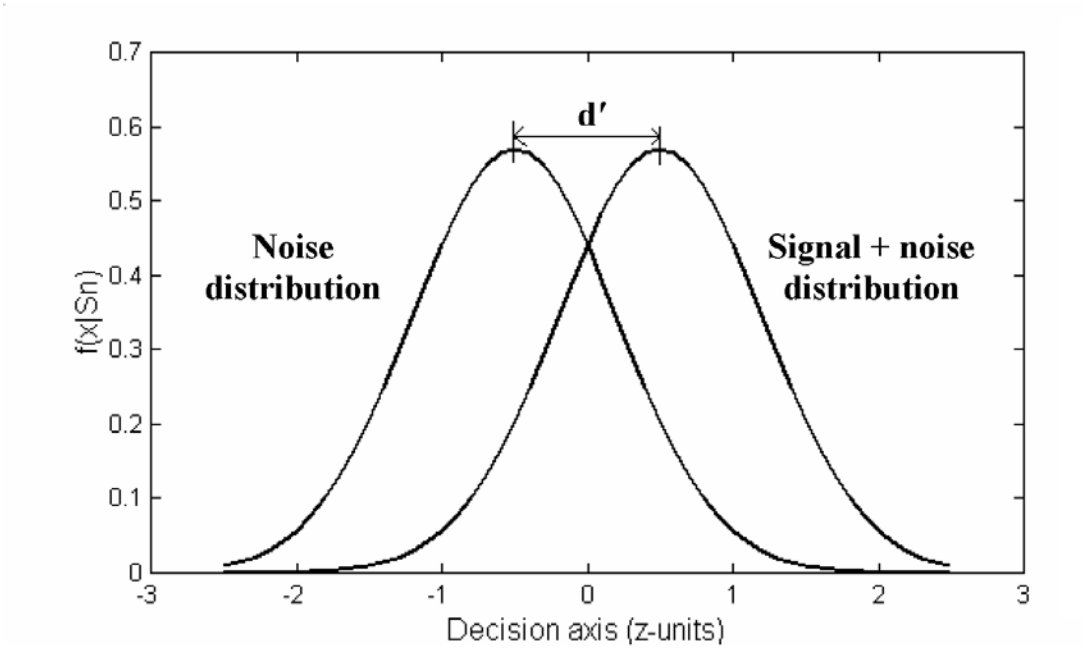

Figure 2.1. The decision space of two stimuli events. The left and right distributions are the internal representations of the noise or lower intensity stimulus and the signal or higher intensity stimulus, respectively. f(x|Sn) denotes the height of the probability density curve of stimulus *n* which represent the likelihood that the participant would choose either N or S for a selected point on the decision (x) axis. The distance between the peaks of the distributions is *d′*. (adapted from Wickens, 2002, p.21)

The Gaussian distribution is widely used within SDT because of the widespread applicability of the distribution to the empirical data (Hanley, 1988). The distribution provides accessibility to the derivation and calculation of results using Gaussian statistics which characterise many modern statistical tests (Simpson & Fitter, 1973). However, other types of distributions other than the Gaussian distribution may also be used to describe the evidence variable (Egan, 1975; Hautus & Irwin, 1992).

In SDT, the discriminability measure is denoted *d′*. It may be defined as the difference between the z-scores of the hit and false alarm rate. Equation 2.3 is the formalisation of this definition (Macmillan & Creelman, 2005, p.8),

$$d' = z(H) - z(F) \tag{2.3}$$

where $z(H)$ and $z(F)$ and are the z-score transformations for the hit rate and false alarm rates respectively.

For z-score transformations, a proportion of 0.5 is converted into a z-score of 0. Proportions larger than 0.5 are transformed into positive z-scores and proportions smaller than 0.5 are transformed into negative z-scores. A conversion table can be referred to for the transformation. This is included in Appendix A. The conversion table utilises the symmetrical property of the z-scores. When two proportions are equally far from 0.5, it leads to the same absolute z-score. For example, proportions of 0.6 and 0.4 are 0.1 units away from 0.5. They have the same absolute z-score of 0.253. After obtaining the z-score magnitude, the valence is added to the z-score. A positive valence is added if *p*, the proportion that has been converted, is greater than 0.5. And a negative valence is added if *p* is less than 0.5, so that

$$z(1-p) = -z(p) \tag{2.4}$$

where *z* is the z-score transformation of the proportion, *p* (Macmillan & Creelman, 2005, p.8).

In Table 2.3, the $H = 0.70$ and $F = 0.40$, so that $z(H) = 0.524$, $z(F) = -0.253$. And using these values, $d' = 0.524 - (-0.253) = 0.777$. Table 2.4 illustrates further the calculation of a series of hit and false alarm pairs.

Hit rate	**False alarm rate**	***z(H)***	***z(F)***	***d'***
0.90	0.60	1.282	0.253	1.029
0.90	0.50	1.282	0.000	1.282
0.95	0.05	1.645	-1.645	3.290
0.55	0.15	0.126	-1.036	1.162
0.60	0.60	0.253	0.253	0.000

Table 2.4. The computation of d' for examples of hit rate, false alarm rate and their associated z-scores

When a participant is unable to discriminate at all, the hit rate is equal to the false alarm rate so that $d' = 0$. For the participant with perfect discrimination ability, the d' is infinite. The largest possible finite d' is dependent on the number of decimal places to which H and F are carried. When H = 0.99, F = 0.01, $d' = 4.65$ and many researchers consider this as an effective ceiling. Moderate performance implies that d' is near unity, which is 1 (Macmillan & Creelman, 2005, p.8).

2.4 Receiver operating characteristic (ROC) curves

A good discriminability measure should be relatively stable when aspects of the study, other than discriminability-related factors, change. Signal detection theory assumes that participants have fixed discriminability when asked to discriminate a specific pair of stimulus classes. However, one aspect of the participant's response that may be variable is their response bias. Response bias is the willingness of the participants to report “yes” as opposed to “no”. If d' is an invariant measure of discriminability, this means that there is more than one possible pair of hit and false alarm rates which give the same discriminability. The only difference between all the possible pairs of hit and false alarm rates is the response bias. For example, a participant whose hit and false alarm rates are 0.80 and 0.40, respectively, can also produce the same discriminability with hit and false alarm rates of 0.60 and 0.20, or 0.35 and 0.07. All of these hit and false alarm rate pairs have a d' of 1.09 and differ only in their response bias.

All possible hit and false alarm rate pairs of the same discriminability can be plotted on a graph of hit rate against false alarm rate to yield a graph called the receiver operating characteristic (ROC) curve (Figure 2.2) (Green & Swets, 1966, p.60; Macmillan & Creelman, 2005, pp.10-11).

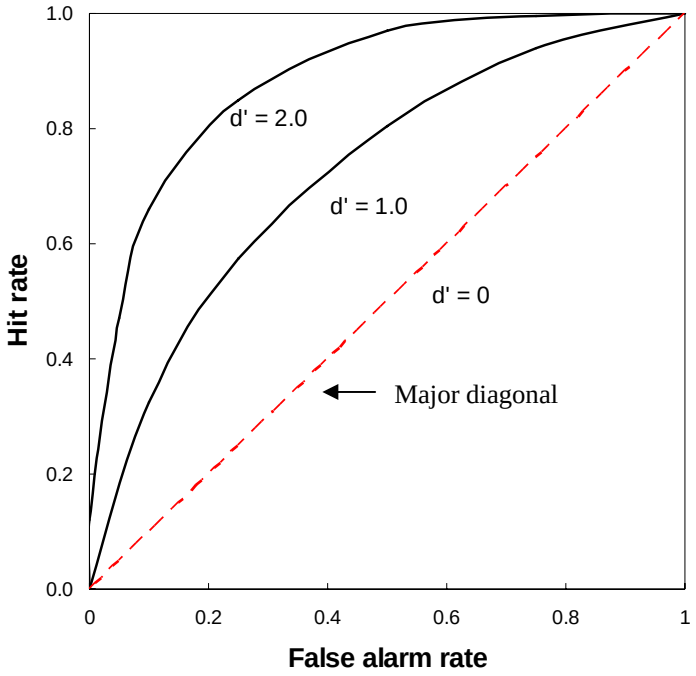

Figure 2.2. Receiver operating characteristics curve (ROC) space. The ROC curves for various *d′* resides in the ROC space. The major diagonal is where *d′* = 0 and discrimination ability is at chance level (adapted from Macmillan and Creelman, 2005, p.10).

The graph is plotted with both hit and false alarm rates having a range of 0 to 1. Therefore, the graph is bounded by the range of proportions of the hit and false alarm rates. The area in which the ROC curve resides is called the ROC space.

When performance is at chance level such that discriminability is zero (*d′* = 0), the ROC is the major diagonal, the diagonal line running from the bottom left corner to the top right corner of the ROC space. At the major diagonal, the hit and false alarm rates are equal. The major diagonal is also called the 'chance line'. As the discriminability of the participant increases, the curve shifts towards the upper left corner of the ROC space.

The theoretical ROC curves in Figure 2.2 illustrate an important characteristic of the theory: the ability to have perfect performance in detecting one type of stimulus will result in the complete failure in detecting the other stimulus type. For example, to have perfect detection of the signal in order to obtain a hit rate of 1, it is necessary

also to have a false alarm rate of 1.This indicates a total failure to correctly reject the noise stimulus. Similarly, to obtain a false alarm rate of 0, a hit rate of 0 is also necessary. These two coordinates are represented by the bottom left and top right corner of the ROC curve (Swets & Pickett, 1982).

The theoretical ROC curve is important in SDT analysis because it is usually fitted to the empirical data to ensure that SDT provides an adequate description of data (Gescheider, 1997, pp.113-116). The theoretical ROC curve can also be used to visually compare the data generated before and after the manipulation of an independent variable.

2.5 ROC curves in transformed coordinates

It is also possible to represent the ROC curves using z-scores to visualise other information about the data. The most important information obtained from ROC curve in transformed coordinates is the slope of the transformed ROC curve. The value of the slope is needed for the determination of the equal variance assumption for the stimuli distributions as depicted in Figure 2.1. The rationale and test of the equal variance assumption will be outlined later in Section 2.8.1.

Equation 2.3, which describes the calculation of d' using $z(H)$ and $z(F)$, can be rearranged to obtain a straight line equation (Macmillan & Creelman, 2005, p11):

$$z(H) = z(F) + d' \tag{2.5}$$

Equation 2.5 describes a transformed ROC function or zROC function (Figure 2.3). Both axes of the ROC curve are expressed in z-scores instead of proportions. After the transformation, the zROC function obtained is a straight line as compared to the convex curve for the ROC curve based on proportions. The range of values for the z-scores is from minus to positive infinity. However, z-scores of 3 and above are seldom encountered. Due to the simple shape of the ROC, a few features can be noted. The y-intercept, that is the value of y when x = 0 in the graph, of the ROC is the d' of the particular participant. And s, the slope of the straight line can be calculated.

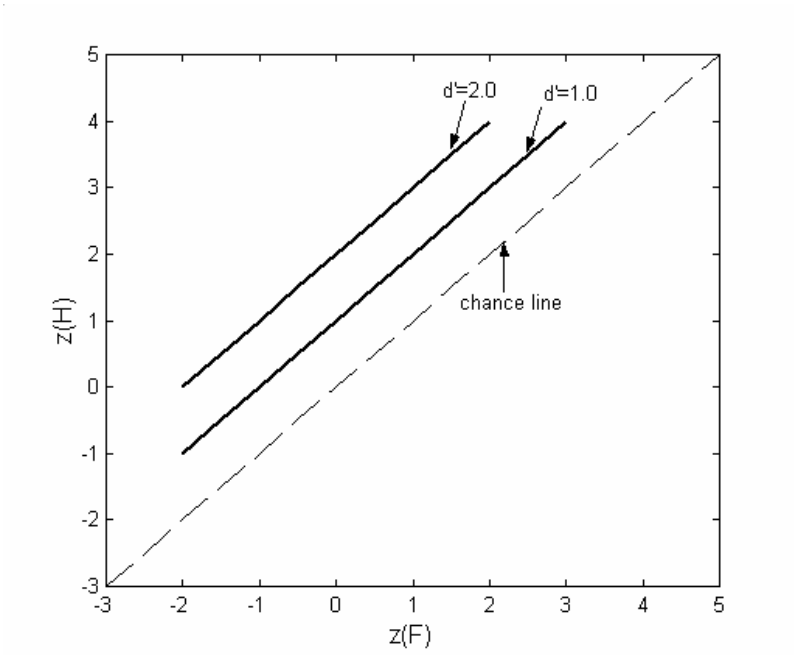

Figure 2.3. Tranformed ROC functions plotted on z-coordinates. z(F) is the transformed false alarm rate in z-units. z(H) is the transformed hit rate in z-units (adapted from Macmillan and Creelman, 2005, p.12).

In detection experiments, researchers aim to achieve a unit slope for the ROC, i.e. the gradient for the slope of the function is 1.0. However, the unit slope is not always experimentally observed. If s is not equal to 1, then d' will not necessarily coincide with the y-intercept. If that situation happens, alternative methods of obtaining d' will have to be used. In Section 2.8 of this thesis, modified measures of discriminability will be explored to accommodate this unit slope departure. This unit slope departure is also related to the testing of the equal variance assumption outlined later in Section 2.8.1.

2.6 Response bias

The response bias in a psychophysical experiment is the tendency of the participant to favour one response over others. This tendency is affected by factors independent of the intensity of the stimuli (Gescheider, 1997, p.404). In the case of the yes-no experiment, it is the tendency to favour either the yes or no response.

The basic response bias measure for SDT is the parameter c (for 'criterion'). It is defined as (Macmillan & Creelman, 2005, p.29):

$$c = -\frac{1}{2}\left[z(H) + z(F)\right] \tag{2.6}$$

When the false alarm and miss rates are equal, the statistic *c* is equal to zero. Negative values occur when the false alarm rate is more than the miss rate, positive values occur when the opposite happens. The extreme values of *c* arise when H and F are both large or both small. For example, if both H and F are 0.99, *c* = -2.33. If both H an F are 0.01, *c* = +2.33.

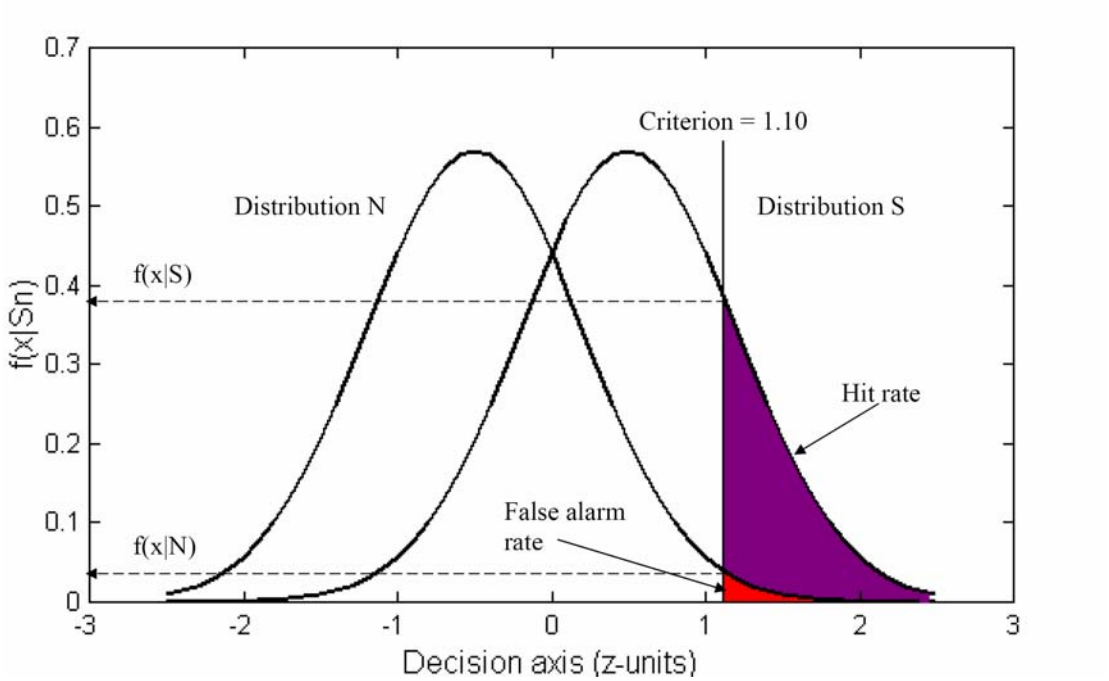

Fig 2.4. Two distributions within a decision space. The criterion is set at 1.10. The purple area indicates the hit rate and the red area indicates the false alarm rate. f(x|Sn) denotes the height of the probability density curve of stimulus *n* which represent the likelihood that the participant would choose either N or S for a selected point on the decision (x) axis.

This is shown in Figure 2.4. The value of *c* is zero when it lies between the intersection of the two probability density function curves. The areas on the left hand side of the criterion signify that the participant will respond ‘no’ to any given stimulus. The areas on the right hand side of the criterion signify that the participant will respond ‘yes’ to any given stimulus. The placement of the criterion will determine how likely the participant will favour a ‘yes’ or ‘no’ response when presented with either the noise or signal-plus-noise stimulus. When a participant’s *c* is high, there is a

tendency to respond ‘no’, and when a participant’s *c* is low, there is a tendency to respond ‘yes’. Figure 2.4 shows that the area lying to the right hand side of the criterion on the left curve is the false alarm rate. And the area lying on the right hand side of the criterion on the right curve is the hit rate.

Another measure for response bias is the likelihood ratio between S and N on any selected point along the decision axis. The relative likelihood of the distribution for the signal stimuli (right curve on Figure 2.4) versus the distribution for the noise stimuli (left curve on Figure 2.4) is called the likelihood ratio. The likelihood ratio is denoted by β. Figure 2.4 shows that for every point on the x-axis, there are two associated likelihoods, one for each distribution. The likelihoods for each distribution in Figure 2.4 are the points where the criterion intersects the curves. And the likelihood ratio is obtained by dividing the two likelihoods:

$$Likelihood\ ratio\,(\beta) = \frac{f(x \mid S)}{f(x \mid N)} \tag{2.7}$$

where $f(x \mid S)$ is the likelihood of the point on the distribution for the signal stimulus, and $f(x \mid N)$ is the corresponding point on the distribution for the noise stimulus (Macmillan & Creelman, 2005, p.33).

The likelihood ratio for the Gaussian distribution can also be derived using response bias measure *c* and discriminability measure *d′* (Macmillan & Creelman, 2005, p.35).

$$\beta = e^{cd'} \tag{2.8}$$

Where *e* is the mathematical constant with the approximate value of 2.718.

This expression can also be manipulated to obtain another commonly used bias measure by taking the logarithm of Equation 2.8, the log-likelihood ratio (Macmillan & Creelman, 2005, p.35):

$$\begin{aligned}\ln\beta &= cd' \\ &= -\frac{1}{2}\left[z(H)+z(F)\right]\left[z(H)-z(F)\right] \\ \ln\beta &= -\frac{1}{2}\left[z(H)^2 - z(F)^2\right]. \end{aligned} \tag{2.9}$$

2.7 The rating experiment

As described above, the results from a yes-no experiment can be used to describe one point on the ROC function. To plot a full ROC function, at least three points have to be obtained. This means that each point should represent a participant with different response biases, but with the proportions to those points describing the same discriminability. There are several strategies to obtain the different response biases from the participant. One of the most efficient strategies is through the use of the rating experiment or rating method.

The rating experiment is very similar to the yes-no experiment. However, in a rating experiment, the participant is given a response set containing more than two categories of responses. Usually the categories are described in terms of the level of confidence that either the signal or the noise was presented to the participant (Figure 2.5). However, it is possible that the categories may also represent the magnitude of the stimulus. For example, in pain perception studies using SDT, the categories usually consist of descriptors of perceived stimulus magnitude (Janal, Glusman, Kuhl & Clark, 1994; Kemperman et al., 1997; Soetanto, Chung & Wong, 2004). Consider a confidence-rating experiment in which the participant is required to discriminate between two stimuli of different intensities, for example two thermal stimuli at 46°C and 47°C. One of the stimuli is randomly presented to the participant. The participant's task is to rate their confidence that the stimuli presented during the observation interval is either the higher intensity stimulus or the lower intensity stimulus. The participant is asked to rate their confidence on a 6-point scale. The participant responds '1' if he/she is absolutely certain the signal presented was of a lower intensity and '6' if he/she is absolutely certain the signal was of a higher intensity. In contrast, the 'yes-no' experiment would force the participants to respond if the stimulus presented is the 'higher' or the 'lower' intensity. Therefore, the

difference between a rating experiment and the yes-no experiment is the use of different response sets.

A.

1	2	3	4	5	6	7	8
Not noticeable	Faintly warm	Moderately warm	Hot, no pain	Fairly painful	Moderately painful	Extremely painful	Withdrawal

B.

1	2	3	4	5	6
I am absolutely certain the weak stimulus was presented	**I am fairly certain the weak stimulus was presented**	**I am somewhat certain that the weak stimulus was presented**	**I am somewhat certain that the strong stimulus was presented**	**I am fairly certain the strong stimulus was presented**	**I am absolutely certain the strong stimulus was presented**

Figure 2.5. A. A magnitude rating scale (adapted from Soetanto, Chung & Wong, 2004). The participant is instructed the sensation elicited by the experimental stimulus on the 8 point scale. B. A confidence-rating scale (Tan, Palmer, Martin & Roche, 2007). The participant is instructed to rate his/her confidence in determining whether the stronger or weaker stimulus was presented during the trial.

2.7.1 Number of categories within a rating scale

The number of categories used is dependent on the objectives of the experimenter. A 6-point scale has often been recommended (Irwin & McCarthy, 1998). There are advantages to the use of a relatively smaller number as opposed to a larger number of categories (sometimes ranging from 12 to100 categories). Participants may find it easier if there were fewer categories within the response set. Research designs in SDT pain perception studies have often used more than 6 categories for their response sets in the study of pain perception, despite the administration of small numbers of trials per stimulus intensity (Clark, Janal, Zeidenberg & Nahas, 1981; Janal et al, 1994). Rollman (1977) has criticised this approach based on the reasoning that unless participants receive a reasonable amount of practice, they will have difficulty in consistently using a large number of categories. There is some evidence to suggest that participants are able to reliably assign stimuli into about three to five categories (Pollack, 1952, Laming, 1984, p.155).

For the purposes of this thesis, a rating response set with six categories is used. This is to facilitate participants in consistently and reliably using a relatively small number of categories for providing responses during the studies.

2.7.2 Analysis of the rating ROC

Once the data has been collected and the results entered into a table indicating the frequencies of response categories corresponding to the appropriate stimulus, a ROC function can be plotted. Table 2.5 is a data set from one participant in one of the studies for this thesis.

Table 2.5A shows the frequency of each rating used when either stimulus 1 (lower intensity) or stimulus 2 (higher intensity) has been presented. And in Table 2.5B these frequencies are converted to proportions. Below the proportions are the cumulated proportions by adding the proportion for particular ratings to the previous rating from rating 6 to rating 1. Rating 1 should display a cumulated proportion of 1 because the maximum probability of a frequency is 1. These cumulated proportions are used to construct the ROC. The two cumulated proportions for each rating are used as a point on the ROC plotted. Using the cumulative proportions for stimulus 1 and 2 for false alarm and hit rates respectively, the data in Table 2.5 would yield the following coordinate pairs: (0.025, 0.500), (0.050, 0.650), (0.200, 0.800), (0.425, 0.950) and (0.775, 0.975). Although there are 6 ratings, only 5 points are produced. This is because the cumulated proportions for rating 1 produce the coordinates of (1,1) and therefore contains no information. Figure 2.6 shows the ROC produced by the information obtained from Table 2.5B.

The assumption behind the use of cumulated proportions to construct the ROC curve is that the participant would possess the same d' even if the stimuli judgments were made with lesser degrees of confidence. In other words, each point represents the varying response biases along the line that constitutes the ROC function. The validity of the rating method has been supported by the finding that the yes-no procedure and rating procedure generally produce similar values of discriminability (Green & Swets, 1966; Markowitz & Swets, 1967).

	Rating					
	1	***2***	***3***	***4***	***5***	***6***
Stimulus 1 (Lower intensity)	9	14	9	6	1	1
Stimulus 2 (Higher intensity)	1	1	6	6	6	20

Table 2.5A. Frequency table of each rating used to judge the stimuli presented for one participant. The participant responses to stimulus 1 and stimulus 2 would be classified as 'false alarm' and 'hits' respectively.

	Rating					
	1	***2***	***3***	***4***	***5***	***6***
Stimulus 1	0.225	0.350	0.225	0.150	0.025	0.025
Cumulative proportions	1.000	0.775	0.425	0.200	0.050	0.025
Stimulus 2	0.025	0.025	0.150	0.150	0.150	0.500
Cumulative proportions	1.000	0.975	0.950	0.800	0.650	0.500

Table 2.5B. Frequencies from Table 2.4 converted to proportions and their respective cumulated proportions for each stimulus. The cumulative proportions from stimulus 1 and stimulus 2 would produce values on the x- and y-axis of the ROC curve respectively.

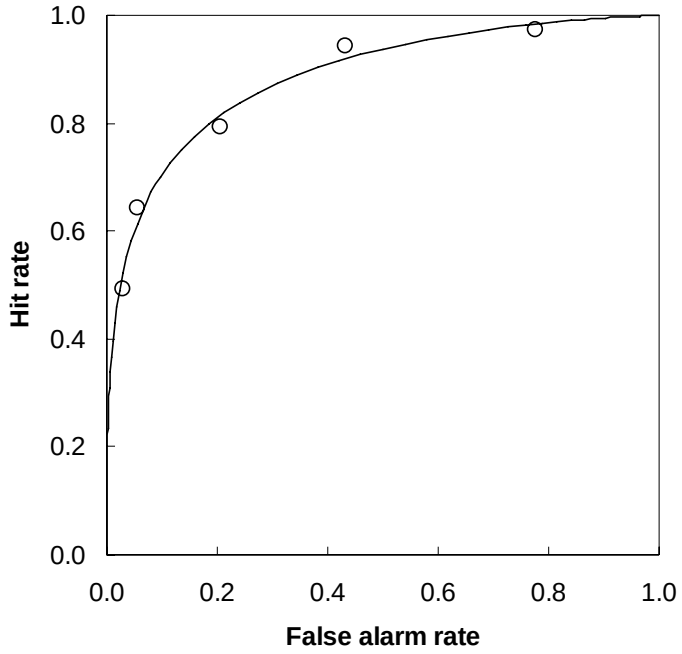

Figure 2.6. An ROC curve constructed through cumulated proportions by using the rating method. Data from Table 2.5 were used to plot the ROC curve.

2.7.3 The one-interval experiment

All the methods described above can be classified under the one-interval experiment. Figure 2.7 summarises the temporal sequence of a one-interval experiment. During the experiment, a pre-selected number of trials, i.e. the number of stimuli-response combination, is administered to the participant. When a trial begins, the experimenter administers one stimulus to the participant. After the stimulus has been administered, the experimenter asks the participant for a judgment. The participant makes a response as to the level of confidence that the either one of the stimuli was presented. Dependent on the experimental setup, a feedback may or may not be given to the participant regarding the status of the stimuli. The trial ends here and the next trial begins.

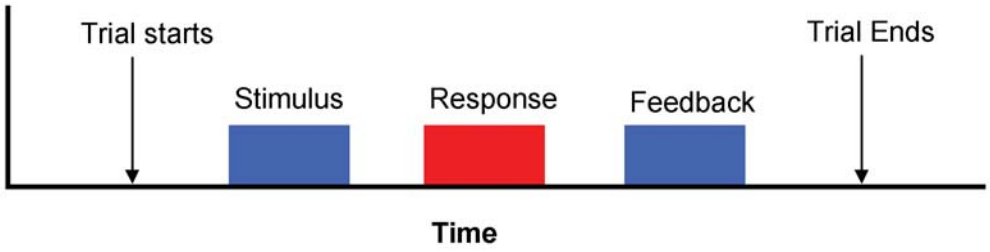

Figure 2.7. Diagrammatic representation for the temporal sequence of a one-interval experiment.

2.8 Unequal-variance Gaussian model of SDT

When *d′* and *c* are used to estimate the discriminability and response bias respectively, an important assumption is made: the signal and noise distributions within the decision space have approximately equal variances. This is called the equal-variance Gaussian model of SDT. Many pain perception studies that used SDT did not adopt the equal-variance model in order to avoid the above assumption (Janal et al, 1994; Naliboff , Cohen, Schandler & Heinrich, 1981; Yang et al., 1985; Lautenbacher, Moltner, Lehmann & Galfe, 1989; Kemperman et al, 1997). Instead, the measures *P(A)* and *B* were chosen to estimate the discriminability and response bias respectively. These measures are deemed to be non-parametric and therefore considered suitable by Clark (1994) for analysing experimental data from pain perception studies. The nonparametric models of SDT will be discussed later in Section 2.9. The following section will describe a method for determining if the data conforms to the equal-variance Gaussian model of SDT. Other models of SDT will also be examined for situations when the data does not meet the assumption of equal-variance.

2.8.1 Determination of the equal variance assumption

The equal-variance assumption can be determined by examining the slope of the transformed ROC function. When the slope of the transformed ROC is 1.0, the data meet the equal-variance assumption. A transformed ROC with a slope of 1.0 is also described as having a unit slope (Macmillan & Creelman, 2005, p.14). The reason behind this is that the y-intercept and the x-intercept of the transformed ROC function represent the hit and false alarm rates in units of standard deviation, i.e. z-scores. Therefore, if we denote the x-intercept to be d'_1 and the y-intercept to be d'_2, the slope

of the transformed ROC function (*s*) is simply the ratio between d'_1 and d'_2 (Macmillan Creelman, 2005, p.59):

$$s = d'_2 / d'_1 \tag{2.10}$$

Therefore, if the slope has a value of 1.0, then the two intercepts must be equal on the transformed ROC function.

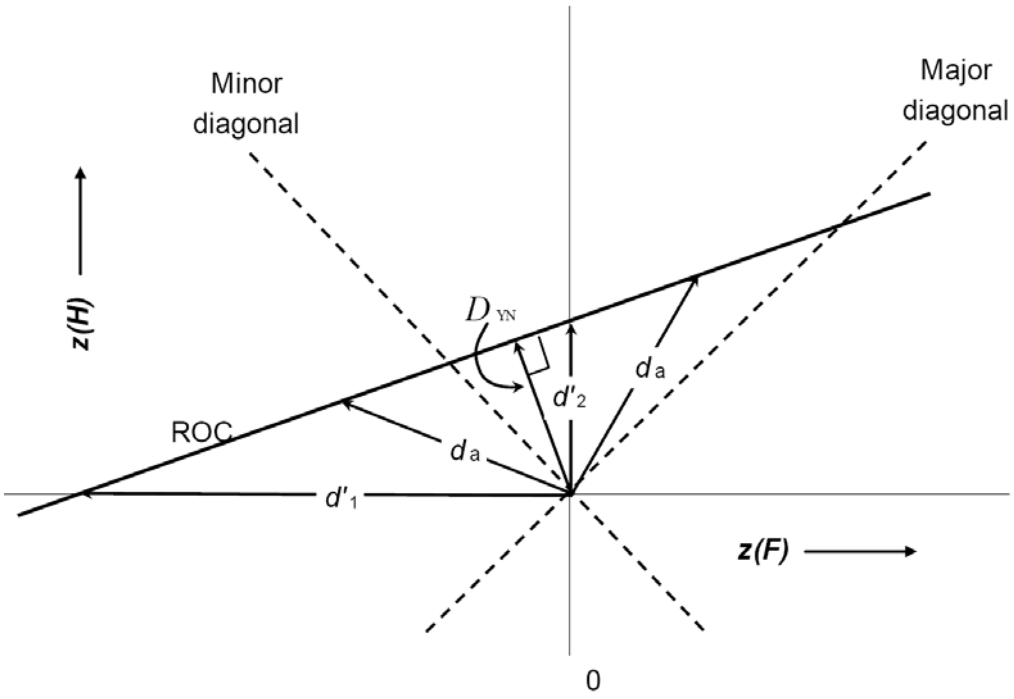

Figure 2.8. Non-unit slope ROC. The figure shows the alternative indices of d_a, d'_1, d'_2, and D_{YN} and their spatial relationship. These indices are explained within the text. The distance of da is $\sqrt{2}$ times as long as D_{YN}. D_{YN} is the perpendicular distance from the origin to the ROC (adapted from Macmillan and Creelman, 2005, p.60).

For the situation when the slope of the function is not equal to 1.0, the y-intercept and x-intercept will not be equal (Figure 2.8). If the transformed ROC function demonstrates this characteristic, this also means that the standard deviations of the signal and noise distributions are not equal. Hence, the equal-variance assumption has not been met and the equal-variance Gaussian model will not fit the data satisfactorily.

2.8.2 Alternative indices for estimating discriminability

There are alternative Gaussian models of SDT that do not require the signal and noise distributions to demonstrate equal-variance. The two most common discriminability measures that adopt an unequal-variance Gaussian model are d_a and d_e (Simpson & Fitter, 1973; Gescheider, 1997, p.129-131; Macmillan & Creelman, 2005, p.62). The choice of the most desirable unequal-variance discriminability index should meet the following criteria: the graphical representation of the index on the transformed ROC space should 1) be between the chance line (the major diagonal) and the transformed ROC function, 2) be shorter than the x-intercept but longer than the y-intercept, and (3) measure the mean distance between the two distributions in unit that is an average of the distributions' standard deviations (Macmillan & Creelman, 2005, p. 61-62). The main reason for using these criteria in determining the most desirable unequal-variance Gaussian discriminability measure is to find a suitable intermediate measure that takes into account the difference between the standard deviations of the two distributions.

Both indices d_a and d_e satisfy all of the above conditions adequately. Macmillan & Creelman suggested that the index d_a , by Simpson & Fitter (1973), is a desirable index not just based on the three criteria, but also because of relative ease in converting d_a to other discriminability measures. This ease in measure conversion is due to the theoretical relationship between d_a and other discriminability measures, in particular the measures d' and A_z.

2.8.3 Computation of d_a

In order to calculate d_a, the following information is required from the transformed ROC space (Figure 2.8): the y-intercept, d'_2, and the slope of the transformed ROC function, s (Equation 2.10). The measure d_a is obtained by dividing d'_2 with the root-mean-square standard deviation, a type of average equal to the square root of the mean of the squares of the standard deviations for the two distributions (Macmillan & Creelman, 2005, p.62):

$$d_a = \frac{d'_2}{[\frac{1}{2}(1+s^2)]^{1/2}} = \left(\frac{2}{1+s^2}\right)^{1/2} d'_2. \qquad (2.11)$$

where d_a is the unequal-variance Gaussian discriminability measure, d'_2 denotes the y-intercept for the transformed ROC function, s denotes the slope of the transformed ROC function and, the root-mean-square standard deviation is $[\frac{1}{2}(1+s^2)]^{1/2}$. Interestingly, when the transformed ROC slope is 1.0, d_a is equivalent to d'. Therefore, if the unequal-variance Gaussian model was used during the design of the experiment, but the slope of the transformed ROC function was determined to be 1.0 based on the collected data, d' may be used instead of d_a. This is because the unit slope of the transformed ROC based on the data has verified that the data may be described by an equal-variance Gaussian model of SDT. Irwin & Whitehead (1991) and Irwin et al. (1994) used this procedure in their experiments where the slope of the transformed ROC slope was established not to be deviating systematically away from s = 1.0. After the equal-variance assumption was confirmed, the equal-variance Gaussian model was adopted. If not, an unequal-variance Gaussian model was chosen instead.

The associated response bias measure for d_a is c_a. c_a is also computed using the root-mean-square standard deviation, $c_a = -\sqrt{2} \cdot s(1+s^2)^{-1/2}(1+s)^{-1}[z(H)+z(F)]$.

In the next section, some of the nonparametric indices commonly used by investigators will be introduced, in particular A' (Pollack & Norman, 1964) and A_g (Pollack & Hsieh, 1969) or *P(A)* (McNicol, 1972), the notation conventionally used in pain SDT studies.

2.9 Nonparametric SDT indices

Over the years, signal detection theorists have devised various indices to suit their experimental requirements and also to explore the theoretical implications of these indices to sensation, perception, cognition, memory and even abstract social constructs. Following the tradition of statistical science, some investigators have developed nonparametric SDT indices to break away from the boundaries of

parametric assumptions imposed on the interpretation of the empirical data. Some might see this development as a continual progression, however, some theorists disagree (Macmillan & Creelman, 2005; Pastore, Carwley, Berens & Skelly, 2003). Table 2.6 shows some of the nonparametric indices that have been developed.

Table 2.6

Nonparametric discriminability and response bias indices

Discriminability index	Formula	Reference
A'	$A' = 0.5 + (H - F)(1 + H - F)/[4H(1 - F)]$	Pollack & Norman (1964); Grier (1971)
Ag or $P(A)$	$A_g = 0.5\Sigma(F_{i+1} - F_i)(H_{i+1} + H_i)$	Pollack & Hsieh (1969)
A_z	$A_z = \Phi(d_a / \sqrt{2})$	
Response bias index		
B'_H	$B'_H = 1 - F(1 - F)/[H(1 - H)]$ *if* $H \le 1 - F$ $B'_H = H(1 - H)/[F(1 - F)] - 1$ *if* $H \ge 1 - F$	Hodos (1970)
B''	$B'' = [H(1 - H) - F(1 - F)]/[H(1 - H) + F(1 - F)] - 1$ *if* $H \ge F$ $B'' = [H(1 - H) - F(1 - F)]/[H(1 - H) + F(1 - F)] - 1$ *if* $H \le F$	Grier (1971)
B	$B = \frac{1 - H_l - F_l}{H_u + F_u - H_l - F_l} \cdot C_l$ *When H+F =1, for the boundary between 2 categories within the rating scale,* *H_u is the hit rate for the upper category,* *F_u is the false alarm for the upper category,* *H_l is the hit rate for the lower category, and* *H_l is the false alarm rate for the lower category.*	McNicol (1972, 2005)

2.9.1 Reasons for usage of nonparametric measures

Macmillan & Creelman (1996) did a sampling of the literature to find out the reasons researchers used nonparametric measures of SDT. The literature sampling protocol they used was to sample 403 papers and list them in chronological order. This list of articles was then used to randomly select one out of a successive set of ten articles. Table 2.7 shows the reasons and the frequency for usage of nonparametric measures within the articles of the sampled list.

Table 2.7 (modified from Macmillan & Creelman, 1996)
Reasons for use of supposedly nonparametric indices

Reasons for use	Frequency
Empirical use	
No distributional assumptions	12
Nonparametric	13
No rationale	3
Passing reference (all but one positive)	6
Theoretical analysis (all but two negative)	6
Total	40

Macmillan & Creelman (1996) reported that 70% of the articles used the nonparametric measures of A' (Pollack & Norman, 1964) and B'' (Hodos, 1970). Macmillan & Creelman (1996) stated that these two measures were popularised by the Grier (1971) paper which continues to be heavily cited to this day. Most of the sampled articles (89%) justified the use of nonparametric indices for their supposedly distribution-free or nonparametric characteristic. According to Macmillan & Creelman (1996), there were a total of six theoretical papers, and four of these were critical of the nonparametric status of these measures. Macmillan & Creelman (1996) also surveyed the areas of research nonparametric measures were used for. These nonparametric measures were mainly used by behavioural scientists in a wide variety of content areas. Table 2.8 shows the areas of study represented in the survey.

Table 2.8 (modified from Macmillan & Creelman, 1996)
Journals in which nonparametric indices were used

Content Area	No. Citations	No. Journals	Journal with Most Citations (No.)
Physiological/animal	141	55	
Animal learning, behaviour	41	9	Journal of the Experimental Analysis of Behavior (21)
Medical	15	11	(four journals with 2 citations each)
Neuroscience	60	28	Neuropsychology (8)
Pharmacological	25	7	Psychopharmacology (13)
Perception/cognition	143	44	
Cognition	26	10	Journal of Memory and Language (7)
Perception	51	12	Perception and Psychophysics (16)
General	66	22	Bulletin of the Psychonomic Society (13)
Other	119	53	
Personality/clinical	41	19	Journal of Abnormal Psychology (6)
Developmental	23	6	Journal of Gerontology (9)
Applied	36	23	Human Factors (8)
Quantitative	19	5	Psychological Bulletin (11)

Macmillan & Creelman (1990) critically explored the character of several supposedly nonparametric measures of SDT. Their conclusion was that most of the nonparametric measures did not warrant the nonparametric classification despite the claims by the authors of these measures. The main reason being that these nonparametric measures contained underlying assumptions similar to those of parametric models (Macmillan & Creelman, 1990; Macmillan & Creelman, 1996, Pastore et al., 2003). The heavily criticised measures were the sensitivity measure A' and the response bias measures B' and B''. Other authors were also in agreement that these supposedly nonparametric measures possess underlying parametric assumptions (McNicol, 1972, p.39; Snodgrass & Corwin, 1988).

2.10 Discriminability indices: A_z and $P(A)$

In this section, the discriminability measures of A_z and $P(A)$ will be considered since these discriminability measures have been either recommended as nonparametric or not criticised as possessing underlying parametric assumptions . The measure A_z will be discussed first as it is a measure favoured by many SDT theorists because of its theoretical elegance and relationship to d_a. The other measure that will be considered is $P(A)$ that was proposed by McNicol (1972) and later used by many researchers in the study of pain perception.

2.10.1 The area under the ROC, A_z

The measure A_z has been noted by Swets (1996) to not assume parametric properties and championed as a possible distribution-free discriminability index. Macmillan & Creelman (1991, 1996) have recommended its use as a truly distribution-free index amongst other “nonparametric” discriminability indices. It is interesting to note that the value of A_z increases consistently with the increase of d'. This signifies that, generally, any form of distribution that can be transformed monotonically to the Gaussian distribution can also be used with the index A_z. This means that A_z can also be used for distributions that have unequal variances (Swets, 1996; p34; Macmillan & Creelman, 2005, p.63). In fact, A_z is related to d_a (see Equation 2.14) (Macmillan & Creelman, 2005, p.63).

2.10.2 Computation of A_z

The procedure to find the area under the ROC in Figure 2.8 is equivalent to finding the value of A_z. To do this, the shortest distance from the origin to the transformed ROC needs to be found. It turns out that Schulman & Mitchell (1966) have already elucidated this measure and it is called D_{YN}. Swets & Pickett (1982) have also named this measure *z(A)*. To obtain A_z, D_{YN} is transformed by the normal distribution function Φ, that is the D_{YN} in z-score units is transformed back into a proportion (Macmillan & Creelman, 2005, p.63).

$$A_z = \Phi\left(D_{YN}\right) \tag{2.12}$$

D_{YN} also happens to be the height of the triangle formed by the hypotenuse length of d_a (side AB of the triangle in Figure 2.9). Isolating the triangle from Figure 2.8 to Figure 2.9, the value of d_a can be found using D_{YN} (Macmillan & Creelman, 2005, p.61).

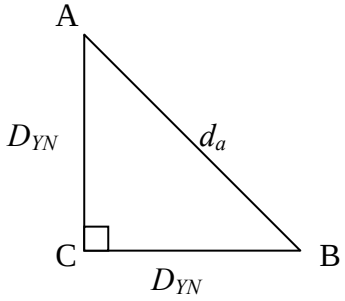

Figure 2.9 Isolation of the right triangle bounded by the opposite and adjacent sides of D_{YN} and the hypotenuse d_a.

To obtain the length AB in Figure 2.9, Pythogoras' theorem is applied,

$$AB^2 = AC^2 + BC^2$$
$$AB = \sqrt{D_{YN}^2 + D_{YN}^2}$$
$$AB = \sqrt{2D_{YN}^2}$$
$$AB = \sqrt{2}D_{YN}$$
$$\therefore d_a = \sqrt{2}D_{YN} \quad (2.13)$$

From Equation 2.13, multiplying D_{YN} by $\sqrt{2}$ will give the value of d_a. This implies that A_z may be found by using d_a. And this relationship is shown below (Macmillan & Creelman, 2005, p.63):

$$A_z = \Phi\left(D_{YN}\right) = \Phi\left(d_a / \sqrt{2}\right) \quad (2.14)$$

As mentioned earlier, A_z was demonstrated to be related to d_a.

The reason A_z is described within this chapter is to illustrate that an area theorem of the parametric models of SDT can be used as a nonparametric application. Several authors have suggested that A_z is a viable nonparametric alternative because it does not require the assumptions of the parameters conforming to normal distributions (Green & Swets, 1966; Swets, 1996).

2.10.3 The area under the ROC, *P(A)*

Another area measure of discriminability is *P(A)* (McNicol, 1972). This measure is also named A_g (Pollack & Hsieh, 1969). *P(A)* is best obtained through the conduct of a rating experiment. In a rating experiment, each category represents a particular point on the ROC. If we plot the points in a ROC space, we can then connect the points with the assumed points of coordinates (0,0) and (1,1) as the extreme ends. Figure 2.10 shows a *P(A)* ROC curve.

We can gain knowledge of the underlying area of the curve by dissecting the graph into trapeziums. Therefore, *P(A)* is simply the addition of the areas formed by the trapeziums.

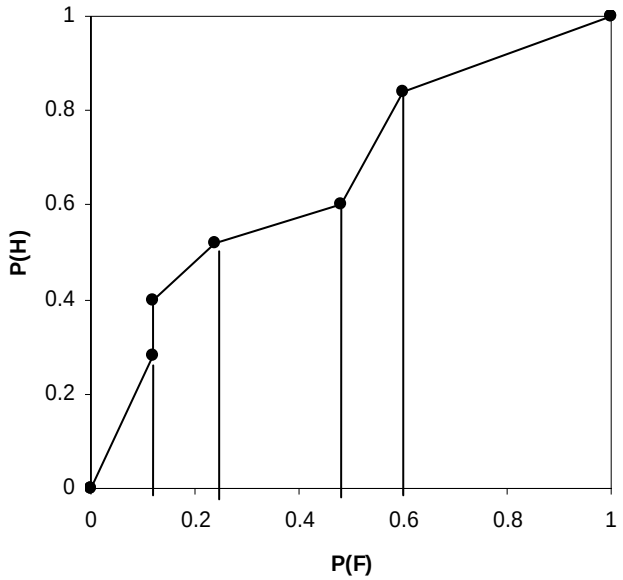

Figure 2.10. A *P(A)* ROC curve. The ROC is formed by "connecting the dots" of each hit-false alarm pair coordinates. The ROC area is calculated by summing up the individual trapezium areas within the ROC.

2.10.4 Justification for the usage of *P(A)* in pain studies

P(A) is an index used in many SDT pain perception studies. *P(A)* was utilised and advocated by Clark's group and they provided several reasons for its use.

The most common reason cited was that the nonparametric outcome makes no assumption about the form of the underlying ROC curve (Clark, 1994). Macmillan & Creelman (2005, p.64) agreed that *P(A)* was a good discriminability measure and that it can be obtained without any model assumptions. It was recommended that *P(A)* is obtained with a large number of ratings, otherwise there is a tendency for *P(A)* to underestimate the true area under the ROC curve. This will be further elaborated when the limitations of *P(A)* are discussed later.

The other reason is that in situations where the number of trials tends to be few, *P(A)* is preferable to the parametric outcomes. The intuitive explanation for this is that fewer numbers of trials will increase the chances of an awkward fit between the parametric SDT model and the data. However, Clark (1994) provided an alternative

explanation. His explanation was that since *P(A)* takes into account many more points than *d′* (the discriminability for the yes-no task) on the ROC, it is therefore less error-prone than *d′*. Clark's explanation is true only if the investigator is experimentally obtaining only one point on the ROC curve. In practice however, when the yes-no task is employed to obtain the discriminability, more than one block of trials are used to produce several points in order to plot the ROC curve. A possible solution to the small number of trials being used in pain perception studies is the combined use of the rating task, a form of discriminability averaging, and an unequal-variance Gaussian model for analysis.

2.10.5 Limitations of *P(A)*

Despite *P(A)* being used as a common discriminability in SDT pain perception studies, Wickens (2002) has proposed two problems associated with the use of this measure.

The first problem relates to the points obtained from a data set to plot the ROC. Many of the data sets do not produce an increasing sequence of points on the ROC. For instance, a non-increasing sequence of ROC points may be encountered when the participants do not use all the rating categories available. This is a common occurrence for studies that use low number of trials (Hautus, 1995). Wickens (2002) argued that the absence of such a sequence will result in the trapezoids not being cleanly drawn. The resulting ROC does not resemble a typical ROC because the slope can proceed from being steep to flat to steep again (Figure 2.10). Wickens (2002, p.70) proceeded to suggest that data smoothing may be required and this may be achieved by fitting the results with a theoretical model. However, this would defeat the purpose of using *P(A)* as a nonparametric measure in the first place. Wickens (2002, p.70) justified this approach by stating that the purpose of the smoothing is to obtain an estimate of what the true operating characteristics might look like. This was most probably performed for the explicit purpose of estimating the amount of deviation inherent in the *P(A)* ROC curve compared to the theoretical ROC curve.

The second problem that Wickens (2002, pp.70-71) identified was the underestimation of the true ROC area by *P(A)*. Assuming that the true ROC is known and this is superimposed onto Figure 2.10 to obtain Figure 2.11. It is immediately apparent that the area under *P(A)* is an underestimation of the true ROC area. This

underestimation of the true area is more severe when the there are fewer points used to plot the ROC or when the points are clustered close together. Figure 2.12 further illustrates the underestimation problem. Figure 2.12A shows the ROC plotted with only two points. When the area under Figure 2.12A is compared to Figure 2.10, the underestimation is shown to be more severe. Figure 2.12B shows the data points are clustered near *P(F)* = 0. The increase in underestimation of the area is evident.

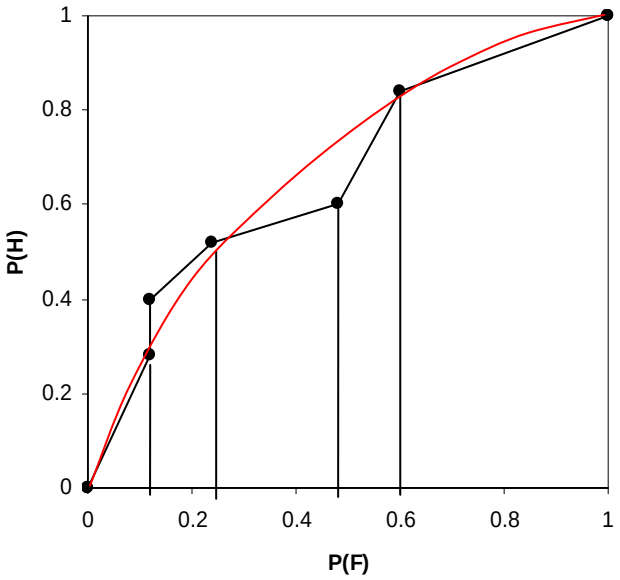

Figure 2.11. *P(A)* ROC with an overlapping true ROC. It is obvious that the *P(A)* underestimates the true ROC area.

2.11 Summary and conclusion

This chapter described the theory and methods underlying signal detection theory. The yes-no and the one-interval rating experiments, with the associated discriminability and response bias measures, were illustrated. The one-interval rating experiment was used in this thesis for collecting participant responses and generating a full ROC curve because it is the most efficient method. A response set consisting of 6 categories was also chosen for the rating experiment used within this thesis. This is to facilitate participants in consistently and reliably providing responses for the study.

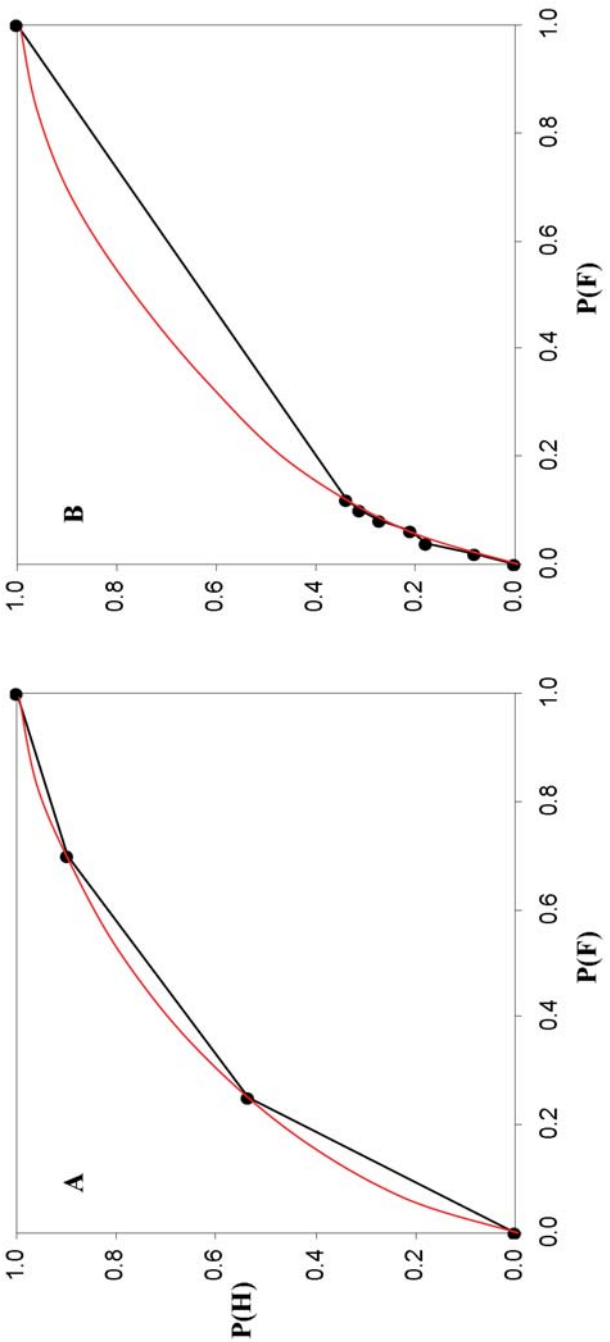

Figure 2.12. Two representations of a similar ROC. A. Only two points are used to obtain the ROC curve. The *P(A)* underestimates the true ROC. B. Six points are used to obtain the ROC curve. The points are clustered near $P(F)=0$, and the underestimation is more acute than curve A (left panel).

Several proposed nonparametric SDT measures have been highlighted. The most common reason for the investigators' choice of these indices is the assumption that these indices are distribution-free. This may not be true for most of the indices. The use of a popular index, *P(A)*, was also considered but this tends to underestimate the true ROC curve. This thesis, therefore, did not use *P(A)* or any of the proposed nonparametric SDT measures. Instead, the parametric measures of d' and c will be used for studies within this thesis. When the slope of the zROC is not equal to 1, the unequal variance SDT index of d_a and c_a will be used instead. The main reason for not using nonparametric indices is that they confer no additional advantages over parametric indices.

Later in this thesis, Chapter 4, Section 4.3 will discuss some issues relating to the use of c when participants do not utilise all of the rating categories available. These issues led to the exclusion of c within this thesis for analysis.

Chapter 3
Signal Detection Theory in the Study of Pain

3.1 Introduction

In the previous chapter, the theory and methods of signal detection theory were described. Some pain perception researchers have used signal detection theory (SDT) for the investigation of nociception and pain perception processes. In particular, SDT pain perception researchers have investigated the influence of depression and anxiety on the SDT measures in both healthy (Dougher, 1979; Malow, 1981) and clinical populations (Malow, West & Sutker, 1989), and examined the effects of antinociceptive agents on SDT measures (Chapman & Feather, 1973; Yang Clark, Ngai, Berkowitz & Spector, 1979; Lineberry & Kulics, 1978; Grilly & Genovese, 1979). Several pain perception studies have also used SDT measures for describing the pain perception characteristics of persons with painful conditions such as chronic low back pain (Naliboff et al., 1981; Cohen, Naliboff, Schandler & Heinrich, 1983; Yang, Richlin, Wagner & Clark, 1985) and irritable bowel syndrome (Dorn et al., 2007).

The following review will focus on chronic low back pain (CLBP) and its associated issues for two reasons. The first reason is that out of the four studies that have investigated painful conditions using SDT methodologies, three studies have focused on the clinical population of CLBP sufferers. The second reason is that the findings for these CLBP studies using SDT methodologies have been consistent. All studies have found that discriminability is lower for CLBP sufferers compared to healthy individuals. Most of these studies have found that the response bias of CLBP sufferers does not significantly differ to healthy individuals.

There are three issues associated with the SDT studies investigating CLBP sufferers. The issues are: 1) the potential influence of psychological factors on the SDT measures, 2) the effect of analgesic medication on pain report by CLBP sufferers as described by SDT measures and, 3) the construct validity of the SDT measures when applied to interpret results of pain perception studies. However, the first two issues are interdependent with the third. Therefore, the first two issues will be examined in

this chapter and Chapter 4 is committed to the third issue concerning the construct validity of SDT measures, the main theme of this thesis. Before the issues are discussed, the SDT studies on CLBP sufferers will be described to provide an overview of the work that has been conducted.

3.1.1 Literature search strategy

The following databases were searched for relevant articles: Ovid MEDLINE (years 1950 to Dec 2007), PsychINFO (years 1906 to Dec 2007) and CINAHL Plus (1937 to Dec 2007). The general search terms used were (signal detection theory$ OR sensory discrimination theory$ OR signal detection$) AND (pain perception$ OR pain$ OR nociception$). The following names of the experts on the topic of signal detection theory in pain perception were also searched (Clark WC$ OR Chapman CR$ OR Rollman GB$). The abstracts from the search results were browsed to locate relevant articles. The reference lists for the articles located were also searched for relevant papers.

It was noted from the literature that most of the research activity within this topic was concentrated between 1950-1990. However, the quantity of publications for this topic fell dramatically after 1990. This decrease in research activity is notable from the dates of publication quoted within this thesis. The reason for this decrease in research activity is unclear. However, it is likely that the critique published by Rollman (1977) may have had a major influence.

3.2 SDT studies on chronic low back pain sufferers

Naliboff et al. (1981) compared the discrimination ability to noxious thermal stimuli of 15 CLBP sufferers with 11 healthy individuals. The noxious experimental stimulation was radiant heat administered via a 250W infrared heat gun. The duration of each noxious stimulation lasted 4 seconds and the body region of stimulation was the participant's forearm. The participant's pain threshold was first determined through an ascending method of limits procedure in order to determine the temperatures used for conduct of the SDT rating experiment. The ascending methods of limits procedure consisted of the heat stimuli increasing in temperature administered on the participant's forearm. The participant was required to indicate

when the thermal stimulus became ‘just detectably warm’ (heat detection threshold) and when the stimulus became ‘faint pain’ (pain detection threshold). The researchers did not report the instructions provided to the participants and the method through which participant responses were given for the method of limits procedure. The following heat gun temperature intensities were then chosen for the SDT procedures specifically for each participant: 1) Blank stimulus (heat gun not switched on), 2) stimulus temperature at heat detection threshold (DT), 3) stimulus temperature at 3°C below the pain threshold temperature (PT – 3°C), 4) stimulus temperature at pain threshold temperature (PT) and, 5) stimulus temperature at 3°C above the pain threshold temperature (PT + 3°C). There were 130 stimulus trials in total. This consisted of 26 trials for each of the five stimulus temperatures. The trials were presented randomly with the experimental constraint that each 5-trial block included each of the stimulus temperatures. For the SDT procedures, the participants rated the intensities on a magnitude-rating scale which consisted of the following six categories: ‘Nothing’, ‘warm’, ‘hot’, ‘faint pain’, ‘moderate pain’ and ‘severe pain’. The nonparametric SDT indices of A' and B'' were computed for the discriminability and response bias indices respectively (Grier, 1971). The computation formula for these SDT measures are reported in Chapter 2, Table 2.6 of this thesis. Naliboff et al. (1981) excluded the stimulus temperatures of ‘blank stimulus’ and DT for the analysis. The researchers found lower discrimination ability of the CLBP sufferers, between the stimulus temperatures of PT -3°C and PT, and this was statistically significant compared to healthy individuals. No statistically significant differences were found for discriminability of the PT and PT + 3°C stimulus pair between groups. No statistically significant difference was found for the response biases of all temperatures between the two groups.

Cohen et al. (1983) compared the discrimination ability to noxious thermal stimuli of 11 CLBP sufferers with 11 age and gender-matched healthy individuals. The experimental set up was the same as Naliboff et al.’s (1981) study, with radiant heat stimuli: of 4 seconds duration to the forearm. The procedure for determining the participant’s pain threshold was also the same. The following heat gun temperature intensities were then chosen specifically for each participant for the SDT procedures: 1) stimulus temperature at pain threshold temperature (PT), 2) stimulus temperature at 3°C above the pain threshold temperature (PT + 3°C), 3) stimulus temperature at 6°C

above the pain threshold temperature (PT + 6°C) and, 4) stimulus temperature at 9°C above the pain threshold temperature (PT + 9°C). There were 104 stimulus trials in total. This consisted of 26 trials for each of the four stimulus temperatures. The trials were presented randomly. For the SDT procedures, the participants rated the intensities on a magnitude-rating scale which consisted of the following nine categories: 'Nothing', 'warm', 'hot', 'very faint pain', 'faint pain', 'mild pain', 'moderate pain', 'strong pain' and 'very strong pain'. The nonparametric SDT indices of *P(A)* and *B''* were computed for the discriminability and response bias indices respectively. The researchers found a lower discrimination ability of the CLBP sufferers, between the stimulus temperatures of PT + 6°C and PT+ 9°C, and this was statistically significant compared to healthy individuals. No statistically significant difference was found for discriminability of the PT and PT + 3°C stimulus pair between the two groups. No statistically significant difference was found for the response biases for all temperatures between the two groups.

Yang et al. (1985) compared the discrimination ability to noxious thermal stimuli of 55 CLBP sufferers with 47 healthy individuals. Information about the type of pain medication consumed for more than two weeks was taken from the patients. This was ranked-ordered on the basis of the medication's strength of effect as determined by Yang et al. (1985): category 1, non-narcotic analgesics; category 2, combined psychotropic and non-narcotic analgesics; category 3, narcotic alone; category 4, combined narcotic and psychotropic. The participants were instructed not to take any pain medication on the day of testing. The CLBP sufferers also rated the intensity of their clinical pain on a 8-cm Visual Analogue Scale (VAS), with the anchors of 'No pain' and 'Pain as bad as it could be' for each end of the scale. The noxious experimental stimulation was radiant heat delivered via a 100W projector lamp bulb heat gun. The duration of each noxious stimulation lasted 3 seconds and the body region of stimulation was on the participants' non-dominant forearm. The following heat gun physical stimulus intensities were chosen for the SDT procedures: 0, 100, 340 and 390 $mcal \cdot cm^{-2} \cdot s^{-1}$ for the CLBP sufferers, and 0, 50, 340 and 390 $mcal \cdot cm^{-2} \cdot s^{-1}$ for the healthy individuals. The researchers' choice of the 50 $mcal \cdot cm^{-2} \cdot s^{-1}$ stimulus instead of 100 $mcal \cdot cm^{-2} \cdot s^{-1}$ for the healthy individuals was because this participant group achieved perfect discrimination (i.e. $P(A) = 1.0$) between the 0-100 $mcal \cdot cm^{-2} \cdot s^{-1}$ comparison during developmental testing. Therefore to increase the difficulty of the

task, a smaller difference in physical stimulus intensity was chosen. There were 32 stimulus trials in total for each participant. This consisted of 8 trials for each of the four stimulus temperatures. The stimulus intensity of the trials were presented randomly. For the SDT procedures, the participants rated the intensities on a magnitude-rating scale which consisted of the following 10 categories: 'Nothing', 'maybe something', 'faintly warm', 'warm', 'hot', 'very hot', 'very faintly painful' 'faintly painful', 'painful' and 'very painful'. Four additional categories were included that described the participants' behaviour if withdrawal from the thermal stimulus was observed: 'withdrawal at the 4th second', 'withdrawal at the 3rd second', 'withdrawal at the 2nd second' and 'withdrawal at the 1st second'. The nonparametric SDT indices of *P(A)* and *B* were computed for the discriminability and response bias indices respectively. The computation formula for these SDT measures are reported in Chapter 2, Table 2.6 of this thesis. The researchers found a lower discrimination ability of the CLBP sufferers, between the stimulus intensities of 340 and 390 mcal·cm^{-2}·s^{-1}, and this was statistically significant compared to healthy individuals. No statistically significant differences were found for discriminability of all the other stimulus intensity pairs between the two participant groups. The researchers also found a statistically significant higher response bias of the CLBP sufferers, between the stimulus intensities of 340 and 390 mcal·cm^{-2}·s^{-1}, compared to healthy individuals. The higher response bias meant that there was an increased tendency for CLBP sufferers to allocate responses to the lower intensity descriptions ratings for the magnitude-rating scale. No statistically significant differences were found for the response biases of all the other stimulus intensity pairs between the two participant groups. There were no statistically significant correlations between the clinical pain intensity (as measured by the VAS) and the type of pain medication consumed for more than two weeks with the discriminability or response bias measures.

3.2.1 Summary

In summary, CLBP sufferers generally have a poorer discriminability to noxious thermal stimuli compared to healthy individuals. Naliboff et al. (1981) and Cohen et al. (1983) found no statistically significant differences in response bias between CLBP sufferers and healthy individuals. In contrast, Yang et al. (1985) found a significantly higher tendency for CLBP sufferers to allocate responses to the lower

rating categories compared to healthy individuals. In other words, the CLBP sufferers were more stoic than the healthy individuals.

There is a consistent pattern in relation to the decreased discriminability of CLBP sufferers as compared to healthy individuals. Naliboff (1981), Cohen et al. (1983) and Yang et al. (1985) did not explore or explain this finding either in their discussion or in any follow up papers. However, it has been suggested that there may be an interaction between psychological factors and pain perception (Fernandez, 2002). It is known that symptoms of depression and anxiety are often reported by CLBP sufferers (Nicholas, Asghari & Blyth, 2008). These psychological factors may potentially alter the discrimination ability of CLBP sufferers. However, the three studies reviewed above did not examine the relationship between psychological factors and the discriminability of CLBP sufferers. It is still possible to gain an insight into the likely relationship between anxiety and depression with the SDT measures (discriminability and response bias) through another route. There are SDT studies that have examined the relationship between anxiety and SDT measures either by experimentally manipulating the level of anxiety in participants (Dougher, 1979; Malow, 1981) or through pseudo-experimental designs (Malow et al., 1989). There are also SDT studies that have examined the relationship between depression and SDT measures using participants with depressive symptoms through pseudo-experimental designs (Dworkin et al., 1995; Kemperman et al., 1997). All these studies have used experimental noxious stimuli for obtaining participant responses for generating the SDT measures. Also, the participant groups within these studies did not suffer from a painful condition. The following sections will review these anxiety and depression-related SDT studies.

3.3 Anxiety, pain and SDT

Dougher (1979) conducted one of the first SDT studies examining anxiety and pain. In the study, healthy college student participants were allocated into two groups exhibiting different anxiety scores (high-anxiety versus low-anxiety). Anxiety level was measured using the Taylor Manifest Anxiety Scale (Hoyt & Magoon, 1954). Forty-eight highest scoring students were included in the high anxiety group and the forty-eight lowest scoring students were included in the low anxiety group. The study

did not describe the proportion of men and women in each group. Each anxiety group was further divided into two groups: one group received experimental instructions that facilitated the participants to report their pain and the other group received inhibitory instructions. For the group receiving facilitative instructions, the participants were advised that a reluctance to report pain is often associated with emotional problems. For the group receiving inhibitive instructions, the participants were advised that a tendency to report pain too quickly is often associated with emotional problems. The exact wording of the instructions was not reported in the paper. For the SDT task, the participants were asked to rate the intensity of two noxious pressure stimuli of different weight. Noxious pressure stimuli were administered using a clamping device (the Forgione-Barber dolorimeter) placed on the fingers of participants. The clamping force can be controlled by placing weights on a lever mechanism attached to the device. The participants were asked to rate the intensity of the pressure based on a 7-point scale: 0 = 'no sensation', 1 = 'slight pressure', 2 = 'moderate pressure', 3 = 'slight discomfort', 5 = 'slight pain' and 6 = 'definite pain' every 10 seconds during the administration of the clamping device. Two weights, one weighing 40g and the other weighing 70g, were placed on the clamping device. Forces of 9.8N and 11.51 N were delivered to the participant's finger when the 40g and 70 g weight were placed on the clamping device respectively. The following pre-determined sequences of either LHHLHL or HLLHLH, where H is the 70g weight and L is the 40g weight, were administered for each participant. The pressure stimulus was removed when either the participant reported definite pain (rating 6) or until 300 seconds elapsed. A total of 6 trials were administered to each participant consisting of 3 trials of 40g and 3 trials of 70g. For the purpose of SDT analysis, the responses for the lower intensity stimulus were arbitrarily defined as ratings obtained after 10 seconds of 40g weight application. Similarly, the responses for the higher intensity stimulus were arbitrarily defined as ratings obtained after 30 seconds of 70g weight application. Responses for ratings 1 to 4 were accumulated to represent 'no pain' responses and responses for ratings 5-6 were accumulated to represent 'painful' responses. Using these arbitrary definitions, the hit rate was the proportion of responses allocated to the 'painful' ratings when administered a 70g weight after 30 seconds. Similarly, the false alarm rate was the proportion of responses allocated to the 'painful' ratings when administered a 40g weight after 10 seconds. The discriminability and response bias measures used were A_g and B'

respectively. The results showed that neither the level of anxiety nor the type of instructions influenced the discriminability of any of the four groups of participants. However, a statistically significant lower response bias was found for the high-anxiety group compared to the low-anxiety group as an ANOVA main effect. The lower response bias for the high-anxiety participant group means that the group tended to allocate their responses to rate the experimental stimuli as 'painful' more than the low anxiety-participants. The results showed that anxiety influenced only the response bias and not the discriminability.

In contrast to Dougher's (1979) study, Malow (1981) found anxiety induced change in the discriminability but not the response bias. Malow (1981) examined the effects of experimentally-induced anxiety on pain perception. Forty-eight students (all men) were recruited as participants for this study. Anxiety was induced by the threat of electric shock via a Grass constant voltage shock generator. Participants were told that when a red light showed on the electric shock apparatus, there was a chance that a shock will be delivered. If a white light was lit instead, a shock would not be delivered. An experimental pressure stimulus was administered using the clamping device similar to Dougher's (1979) set-up. For the SDT task, the participants were asked to rate the intensity for two noxious pressure stimuli of different weight. One weight was 40g and the other weight was 70g. The participants were asked to rate the intensity of the pressure based on a 13-point scale: 0 = 'no sensation' to 12 = 'definite pain' every 5 seconds during the administration of the clamping device. The sequence of weight administration or randomisation was not reported in the paper. The pressure stimulus was removed when either the participant reported pain (rating 9) or 60 seconds elapsed. A total of 6 trials were administered to each participant consisting of 3 trials of 40g and 3 trials of 70g. For the purpose of SDT analysis, the responses for the lower intensity stimulus were arbitrarily defined as ratings obtained after 10 seconds of 40g weight application. Similarly, the responses for the higher intensity stimulus were arbitrarily defined as ratings obtained after 30 seconds of 70g weight application. Malow (1981) defined the hit rate as the proportion of responses allocated to ratings 10 to 12 when administered a 70g weight after 30 seconds on 73% of the time. Similarly, the false alarm rate was the proportion of responses allocated to the ratings 10-12 when administered a 40g weight after 10 seconds on 12% of the time. The discriminability and response bias measures used were A_g and B' respectively.

The study showed that the participants had lower discriminability during the red light condition (indicating induced anxiety) and this was statistically significant as compared to the white light condition (indicating no anxiety). The response bias did not show statistically significant results between the two conditions.

Malow (1981) did not provide an explanatory model for their findings. However, Arntz, Dreesen & Merckelbach (1991) have suggested that attention may modulate the relationship between anxiety and pain. The threat of an electric shock delivered during the red light condition may have diverted the participant's attention from the task of discriminating between the pressure pain stimuli. This led the participant to perform poorer in the experimental task, hence the lower discriminability. The attentional modulation of nociceptive neural activation has been shown to be mediated at the medullary dorsal horn (Bushnell, Duncan, Dubner & He, 1984) and medial thalamus (Bushnell & Duncan, 1989). This electrophysiological evidence suggested that attentional modulation of pain may preferentially involve the pathways through the medial thalamus to the anterior cingulate cortex (Villemure & Bushnell, 2002). This pathway has been shown to be involved in pain affect in humans (Rainville et al., 1997). Hence, the advancement of the argument that attention may be a modulating factor for affect in pain perception processes.

This thesis interpreted the results of Dougher's (1979) and Malow's (1981) studies to suggest that anxiety decreases discriminability through central nervous system activity modulation via attentional processes (Villemure & Bushnell, 2002). It could therefore be predicted that a decrease in anxiety level will lead to an increase in discriminability. Malow et al. (1989) examined this hypothesis in highly anxious detoxified substance abusers. Two hundred and twenty patients admitted to a drug detoxification programme were requested to complete the State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, Lushene & Vagg, 1983) to determine their anxiety levels. Sixty-two patients met the criteria for high anxiety. Malow et al. (1989) specified a STAI-state anxiety cut-off score of 50 to be defined as high anxiety. An experimental pressure stimulus was administered to the patients via the clamping device described in Dougher's (1979) study to obtain baseline discriminability and response bias. The SDT test procedure was the same as Malow's (1981) study, including the type and number of ratings, and the number of trials administered. The

discriminability and response bias measures used were A_g and B' respectively. After the detoxification programme, the patients were again administered the STAI to identify those who had a substantially reduced anxiety state (n = 15) and those whose anxiety state remained elevated (n = 15). The baseline discriminability and response bias were not statistically significantly different between the group with post-programme reduced anxiety and the group with post-programme elevated anxiety. The SDT task using experimental pressure stimuli was presented to the patients again. The results showed that the patients with reduced anxiety levels had statistically significantly higher discriminability post-programme compared to baseline. Those patients whose anxiety levels did not improve had a statistical significantly lower discriminability post-programme compared to baseline. The post-programme discriminability for the improved patients was statistical significantly higher compared to the unimproved patients. This provided some evidence that anxiety is associated with lower discriminability to a noxious discrimination task. Patients with reduced post-programme anxiety also showed a statistically significantly higher response bias compared to pre-programme. This meant that the patients with higher post-programme anxiety were less likely to rate the experimental pressure stimulus as 'painful' during the post-programme testing session.

Dougher (1979), Malow (1981) and Malow et al. (1989) used only a single pair of stimuli in their studies for generation of the SDT data. Schumacher & Velden (1984) commented that the choice of only one stimulus pair provides information only specifically to that stimulus pair regarding the participants' discriminability. They argued that a range of stimulus intensities was required instead. Schumacher & Velden (1984) investigated the effect of experimentally-induced anxiety using electric shock, on discriminability of five pairs of electrical stimuli. The electric shocks were induced using a battery-supplied electronic shocker. The researchers did not describe which region of the body the electric shocker was administered. Anxiety was measured by the State-Trait Anxiety Inventory (STAI) at the beginning and the end of the SDT procedures. The researchers did not report the intensity of currents used for the electrical stimuli. Each experimental stimulus lasted 0.5 second. There were a total of 400 trials administered, 80 trials per stimulus pair (40 trials for the lower intensity and 40 trials for the higher intensity) for 5 stimuli pairs. Six students (women) were recruited as participants for this study. The participant responses for

the SDT rating task occurred in two stages. The first stage required the participant to judge if the lower or higher of two stimuli intensities were administered. The second stage required the participant to judge the confidence of the former decision on a four category confidence scale: ‘very certain’ to ‘very uncertain’. The discriminability measure used was d'_s. This index is similar to d_a but instead of the root-mean-squared standard deviation of the stimulus distributions used for computation (Chapter 2, equation 2.11), an arithmetic mean standard deviation is used. The investigators did not report their response bias results because it was argued it was not directly related to the construct of nociception and pain perception. It was the discriminability result that was more important in this particular research context. The investigators found a statistically significant decrease in discriminability for the anxiety phase compared to the baseline (no anxiety) for the two lowest intensities of electrical stimuli. Schumacher & Velden (1984) attributed the decrease in discriminability for the two lowest intensities of electrical stimuli to the modulating effects of attention on the relationship between anxiety and pain. This was similar to the mechanism hypothesised by Malow et al. (1989) and Arntz et al. (1991). The investigators also found a statistically significant increase in the discriminability for the anxiety phase compared to the pre-anxiety phase for the three highest intensities of electrical stimuli (i.e. the opposite of what was observed for the lower intensities). Schumacher & Velden (1984) did not offer an explanation for the higher discriminabilities found for the higher intensities of electrical stimuli.

Nevertheless, it is possible that attention is still the modulating factor between anxiety and pain. As the intensities of the stimuli pairs increased, the quality of experience for the electrical stimuli may have become more similar to the electric shock threat. Instead of attending to the potential threat of electric shock in the high intensity conditions, the participants may have attended to the stimuli pairs as if they were the anxiety-causing shocks. Therefore increased attention towards the higher intensity electric stimuli could have caused the increased discriminability. However, this is only speculation. Studies that have used different modalities for inducing experimental anxiety and the experimental task usually find that the experimental task performance was worsened in the anxiety condition (Malow, 1981; Malow et al., 1989; Peters, Vlaeyen & Kunnen, 2002). These findings raised an associated issue. The context of the study could have an influence on the focus of the attentional

mechanisms used in the discrimination task. When this contextual issue is translated to a clinical pain study setting, the procedure of experimental noxious stimulus compared with the clinical pain could yield interesting interactions. Depending on the perceived noxiousness of the clinical pain compared to experimental noxious stimuli, the participants' discriminability could move in either direction depending on which captures more attention.

3.4 Depression, pain and SDT

Dworkin et al. (1995) examined SDT measures of discriminability and response bias in patients with depression. In their study, three groups of participants were compared: patients diagnosed with major depression (N=26), bipolar disorder (N=15) and healthy individuals without depression (N=32). The patients with major depression and bipolar disorder were diagnosed on the basis of the criteria of the Diagnostic & Statistical Manual of Mental Disorder III (revised) (DSM-III-R) (American Psychiatric Association, 1987). Two pairs of radiant thermal stimuli were administered to the participants' non-dominant forearm for discrimination in the SDT procedure via a projector bulb heat gun. The lower intensity thermal stimuli were 0 and 75 $mcal \cdot s^{-1} \cdot cm^{-2}$ and represented intensities experienced in the "maybe-something" to "warmth" categories used in the response set by the investigators. The higher intensity thermal stimuli were 270 and 300 $mcal \cdot s^{-1} \cdot cm^{-2}$ and represented intensities experienced in the "very hot" to "painful" categories. Each stimulus duration lasted for 3 seconds. Participants were required to rate their responses on a 12-point rating scale: 'nothing', 'maybe something', 'faintly warm', 'warm', 'hot', 'very hot' , 'very faint pain', 'faint pain', 'painful', 'very painful' and, 'excruciatingly painful'. If the participant withdrew his/her forearm before the stimulus duration was completed, the trial was rated as 'withdrawal' by the researcher. The discriminability and response bias measures of d' and $\ln\beta$ (Chapter 2, equation 2.9) respectively, were computed from the data. No statistically significant difference in discriminability was found between the major depression, bipolar disorder groups and control groups for the higher intensity stimuli. The exception was the statistically significantly lower discriminability score between the major depression and control groups for the higher intensity stimuli pair. No significant difference was found for the low intensity stimuli discriminability scores between any groups. For the response bias results, the major

depression group had statistically significant higher response biases compared to the control group for the lower intensity stimulus pair. Also, the bipolar disorder had statistically significant higher response biases compared to the control group for the lower intensity stimulus pair. For the high intensity stimuli, only the major depression group showed a statistically significant higher response bias compared to the control group. The higher response bias meant that participants were more likely to allocate their responses to the ratings in the lower sensory intensity descriptions, i.e. less likely to report the stimuli as painful. Dworkin et al. (1995) suggested that the lowered discriminability for the higher intensity stimuli in the major depression group was not due to cognitive processes, such as poor attention. In other words, if attention was influencing the participants' discriminability, then this would be a global process reflecting in lower discriminability for both the higher and lower intensity stimuli pair. The investigators concluded that depression may have an effect on the SDT measures by decreasing discriminability and increasing response bias.

Kemperman et al. (1997) conducted a study involving women with borderline personality disorder (BPD) and self-injury behaviour (SIB). Some self injurious patients report experiencing no pain during the self-injurious behaviour. These patients also report less pain and an enhancement of mood during a laboratory pain procedure (Russ et al. 1992, Russ, Shearin, Clarkin, Harrison, & Hull, 1993). Kemperman et al. (1997) compared four participant groups in her study. BPD patients who reported pain during SIB (BPD-P) (n = 17), BPD patients who did not report pain during SIB (BPD-NP) (n = 9), BPD patients who did not have SIB (BPD-C) (n =8), and healthy individuals (n = 7). The participants completed the Sheehan Anxiety Scale (Sheehan, Raj, Sheehan & Soto, 1988) and the Beck Depression Inventory (Steer, Clark, Beck & Ranieri, 1999) to obtain data regarding their levels of anxiety and depression respectively. The medication consumption of the participants was also obtained and categorised under these groupings: 'antidepressants', 'antipsychotics', 'mood stabilisers' and, 'benzodiazepines'. For the SDT procedure, noxious thermal stimuli were inflicted on the participants using a 100W projector bulb heat gun on the both the participants' forearms. Four thermal stimulus intensities were used: 50, 100, 320 and 370 $mcal \cdot s^{-1} \cdot cm^{-2}$. Each stimulus lasted for the duration of 3 seconds. After the stimulus had been administered, participants were required to rate their responses on a 8 point scale: 'nothing', 'maybe something', 'warm', 'hot but not painful', 'faint

pain', 'moderate pain' and, 'severe pain'. Similar to Kemperman et al.'s (1997) study, if the participant withdrew his/her forearm before the stimulus duration had been completed, the trial was rated as 'withdrawal' by the researcher. The SDT measures of *P(A)* and *B* were computed. The study found that the BPD-NP group had statistically significant lower discriminability compared to the other participant groups. Also, the BP-NP had statistically significant higher response bias when compared to the BPD-C group. The higher response bias meant that participants were less likely to report the stimuli as painful. The amount of medication between the patient groups was not statistically significant. A correlational analysis was performed to establish the strength of association between the anxiety and depression scores with the SDT measures. In contrast to previous results, anxiety and depression scores were not associated with discriminability and response bias for any of the BPD groups.

The SDT anxiety and depression studies reviewed did not explicitly investigate the mechanisms involved in changes to the SDT indices associated with the psychological variables. Nevertheless, the attentional process has been proposed to be a potential modulating factor between anxiety and noxious discriminability and response bias (Arntz et al., 1991). All the SDT anxiety and depression studies were conducted on participants without painful conditions. Therefore these findings may not be directly generalised to people with painful conditions. In other words, studies using SDT methodology conducted on CLBP sufferers should include the psychological factors of anxiety and depression either as independent variables or as covariates in the study design. The study, in Chapter 8, of this thesis has collected information about the level of depression and anxiety in order to correlate them with the SDT measures.

3.4.1 Summary

In general, the anxiety studies using SDT methodology found that the presence of experimentally induced anxiety produced a lower discriminability or lower response bias. This means that the participants had a poorer ability to differentiate between the experimental stimuli presented to them (lower discriminability) or are more likely to rate the experimental stimuli as 'painful' (lower response bias). For the depression studies using SDT methodology, it was generally found that the presence of depressive symptoms produced a lower discriminability as well as a higher response bias. This means that the participants with depressive symptoms are poorer at

differentiating between the experimental stimuli and are less likely to rate the experimental stimuli as 'painful'. Since anxiety and depression may influence the discriminability and response bias of participants, it is recommended that future studies conducted on populations with depression or anxiety comorbidities should collect the associated data for establishing the covariance. The study in Chapter 8 of this thesis investigated the construct validity of discriminability as an indicator of psychological state (i.e. depression and anxiety). Information about the level of depression and anxiety were collected for the sample of CLBP sufferers. The level of depression and anxiety was then correlated with the SDT measures used to establish their strength of association.

3.5 Pharmacological anti-nociception and SDT

Chronic pain sufferers are often prescribed multiple medications for the symptoms associated with their pain. These pharmacological products fall under several categories. The categories include, but are not restricted to, non-steroidal anti-inflammatories (NSAIDs), aspirin/paracetamol, antidepressants, anxiolytics, muscle relaxants, benzodiazepines, barbiturates and narcotics. These medications may have a primary or secondary analgesic effect. This raises the question of whether the consumption of these medications will have an effect on the SDT measures in clinical studies. From the three CLBP studies using SDT methodology reviewed earlier in Section 3.2, only Yang et al. (1985) has collected data on the medication history of the participants. The following section will review some of the studies examining the effect of analgesic medication on changes in the SDT measures. Most of these studies have focused on the following pharmacological categories: anxiolytics and opioids. Therefore, these pharmacological categories will also be the focus of this review.

3.5.1 Effects of diazepam and morphine on pain perception

One of the first studies to examine the effects of diazepam, a benzodiazepine-derived anxiolytic, on pain perception using SDT procedures was conducted by Chapman & Feather (1973). The aim of the study was to examine if 10mg of orally administered diazepam decreased the participants' pain report as compared to placebo when SDT was used as the analytical model. A total of 30 participants were recruited for this study. A within-subject design was used with the type of medication (diazepam versus

placebo) as one within-subject independent variable and the priority of medication administration (diazepam first or placebo first) as the second within-subject independent variable. The pain threshold was determined by the Method of Constant Stimuli. However the researchers did not describe how this psychophysical method was conducted for their study. Noxious radiant heat was delivered via a projector bulb heat gun on the participants' forearm. Each stimulus lasted 3 seconds. Five different intensities were chosen for the experimental stimuli: The first thermal intensity was zero (that is, the dolorimeter was not switched on), the third intensity was equivalent to pain threshold. The second intensity was 30mcal·cm^{-2}·s^{-1} below threshold and the fourth and fifth intensities were both above pain threshold separated by 30 and 60 mcal·cm^{-2}·s^{-1} respectively. Participants were required to rate their responses on a 5-point scale: 'nothing', 'heat', 'faint pain', 'moderate pain' and 'strong pain'. The SDT measures of d' and c were used for discriminability and response bias respectively. The results showed that there were no statistically significant differences between the discriminability scores of participants who were given diazepam and those that were given placebos. There was also no statistically significant difference in the response bias between the diazepam and placebo groups. With this result, the authors concluded that diazepam did not alter the sensory aspect of the pain report or the participants' willingness to report pain, i.e. the response bias.

Yang et al., (1979) also investigated the analgesic effects of diazepam in addition to morphine, an opioid, using SDT methodology. Twenty healthy men were recruited for the study and were randomly allocated to three intervention groups: 1) 0.14mg/kg diazepam, 2) 0.14mg/kg morphine and, 3) 10 ml saline solution. The medications were administered intravenously. This is in contrast to Chapman & Feather's (1973) study whereby the medication was administered orally. Noxious radiant heat was delivered via a projector 100W bulb heat gun on the participants' forearm. The entire test session consisted of 8 periods, one pre-injection period and seven post-injection periods of 15, 45, 75, 105, 135, 165, 195 min. A total of 98 stimuli trials were administered within each period, 14 trials at each of the 7 thermal intensities: 0, 90, 180, 270, 320, 370 and 400 mcal·cm^{-2}·s^{-1}. The participants were required to rate their responses on a 10-point scale: 'nothing', 'maybe something', 'faint warmth', 'warm', 'hot', 'very hot', 'very faint pain', 'faint pain', 'pain', 'very painful'. If the participant withdrew his/her forearm before the stimulus duration has been completed, the trial

was rated as 'withdrawal' by the researcher. There are four additional categories of withdrawal on the rating scale based on time before withdrawal: 1) 2.66 to 3.00 seconds, 2) 2.32 to 2.65 seconds, 3) 2.00 to 2.31 seconds and, 4) less than 1.99 seconds. The SDT measures of *P(A)* and *B* were computed for the discriminability and response bias of the participants respectively. Results of the study[1] showed that the diazepam and morphine did not significantly decrease discriminability for the noxious stimuli intensity pairs of 320-370 mcal·cm^{-2}·s^{-1} and 370-400 mcal·cm^{-2}·s^{-1} when compared with the saline groups. However, for the lower stimulus intensities (less than 320 mcal·cm^{-2}·s^{-1}), both diazepam and morphine did decrease the participants' discriminability as compared to the saline group. The morphine group had a statistically significant increase in response bias as compared to baseline and the saline group for the higher stimulus intensities (equal or higher than 320 mcal·cm^{-2}·s^{-1}) but not for the lower stimulus intensities. The diazepam group had a statistically significant increase in response bias as compared to baseline and the saline group for both the higher and lower stimulus intensities.

In an animal study by Lineberry & Kulics (1978), the investigators compared the effects of diazepam, morphine and saline administered intramuscularly in rhesus monkeys. A total of 12 male rhesus monkeys were used in this study. Noxious electrocutaneous stimuli were administered using a constant current stimulator. The following 5 electrical stimulus intensities were used: 5, 10, 15, 20 and 25 mA. A total of 200 trials were presented for each session, which consisted of 40 trials per stimulus intensity. The monkeys were trained on a behavioural version of the yes-no task. A yes response was defined as pressing a bar placed in front of the monkey within 1.5 seconds after the administration of a trial. A no response was defined as pressing the bar after 1.5 seconds or if no response was given. The SDT measures of d' and β were computed for the discriminability and response bias respectively. The results showed that there was no statistically significant change in discriminability for either the diazepam or morphine conditions compared to the saline group. There was a statistically significant increase in the response bias for the diazepam condition compared to the saline condition. There was no statistically significant change in response bias for the morphine condition compared to the saline condition. The

[1] NAPS Document No. 03413 c/o Microfiche Publications P.O. Box 3515, Grand Central Station, New York, NY 10163-3513.

discriminability and response bias between the diazepam and morphine groups were not directly compared. No numerical figures were provided for inspection of the descriptive data. All analyses were graphically-based, comparing ROC curves of the drug group and saline.

Grilly & Genovese (1979) obtained contrasting results in their SDT pain studies using rats. Five male Sprague-Dawley rats were used for this study. The rats were trained to either turn left or right in a maze when the appropriate stimulus was presented. The directions within the maze had been paired with the stimuli being discriminated. Reward reinforcement was used to train the rats before the commencement of the actual experiment. Noxious electrical stimulation was administered via a constant current A.C. shock generator. The intensities of the electrical stimuli were 0.075 (non-noxious) and 0.25 mA (noxious). A total of 180 trials were administered (90 trials per stimulus intensity). The duration of both electrical stimuli was 0.1 second. Twenty minutes prior to actual testing, the rats were injected interperitoneally with either morphine ($2mg \cdot kg^{-1}$ or $4mg \cdot kg^{-1}$ morphine sulphate solution) or saline solution. The investigators found that the morphine significantly reduced discriminability but did not affect response bias as compared to the saline condition. It was also found that the lower discriminability was statistically significant for the $4mg \cdot kg^{-1}$ morphine sulphate condition as compared to the saline condition. However, the lower discriminability was not statistically significant for the $2mg \cdot kg^{-1}$ morphine sulphate condition when compared to the saline condition.

3.5.2 Summary

The human studies reviewed above generally showed that diazepam and morphine did not alter the discriminability to experimental stimuli in the noxious range of stimulus intensities. However, the effect is manifested as an increase in the participants' response bias. This result could be interpreted in two ways. The first interpretation is that diazepam and morphine did not produce an analgesic effect in the studies by Chapman & Feather (1973) and Yang et al., (1979). However, both diazepam and morphine decreased discriminability for the non-noxious stimuli intensities (Yang et al., 1979). Also, the animal studies by Grilly & Genovese (1979) showed a dose-response relationship between morphine dosage and discriminability. This provides some evidence that morphine and diazepam most likely produced an analgesic effect.

The second interpretation is that SDT measures are not sensitive to the changes induced by diazepam and morphine, as shown in the results by Chapman & Feather (1973) and Yang et al., (1979). This is a more complex issue involving several domains of the construct validity of SDT measures for nociceptive and pain perception processes. This issue will be discussed in Chapter 4.

The above studies have only investigated the effects of two specific pharmacological compounds that CLBP sufferers may be consuming for their pain or associated symptoms of affective disorder. Polypharmacy, the consumption of several classes of medication, is common for CLBP sufferers (Masters et al., 1992). In order to account for the effects of other medication classes that have not been examined in detail, the association between medication consumption and the SDT measures should be included in the designs of future clinical research. For the study in Chapter 8 of this thesis, a method known as the Medication Quantification Scale is used to address this issue.

3.6 Conclusion

This chapter described three SDT studies on CLBP sufferers. Three issues associated with the studies were identified. The first issue is the potential influence of psychological factors on the SDT measures. Studies on healthy individuals and participants with non-painful conditions showed that the presence of anxiety generally decreased participants' discriminability to noxious stimuli and increased the tendency to label the noxious stimuli as 'painful'. These studies also showed that depression generally decreased participants' discriminability to noxious stimuli and decreased the tendency to label the noxious stimuli as 'painful'. It is recommended here that information about depression and anxiety are collected for future studies conducted on painful conditions. The second issue concerns the effect of analgesic medication on pain report by CLBP sufferers as described by SDT measures. SDT studies investigating the effects of diazepam and morphine in humans showed that these pharmacological compounds do not significantly alter discriminability for noxious stimuli. However, both compounds did increase the participants' tendency to report the noxious stimuli as 'painful'. No study using SDT methodology has so far examined the association between the amount of medication consumed and the SDT

measures on clinical populations with painful conditions. Following up on the above two issues raised, data were collected on the level of anxiety and depression in CLBP sufferers and the quantity of different medication that they consumed for the study in Chapter 8 of this thesis. The third issue concerning the construct validity of the SDT measures when applied to interpreting results of pain perception studies will be discussed in the next chapter.

Chapter 4
Construct Validity of SDT in Pain Research

4.1 Introduction

The strength of SDT in determining the response bias from the sensory/ perceptual variable in question is notable. In the previous chapter, some of the research using SDT analysis for the study of pain perception was reviewed. Two issues of 1) the potential influence of psychological factors on the SDT measures and, 2) the effect of analgesic medication on pain report by CLBP sufferers as described by SDT measures have also been outlined. This chapter will focus on the third issue regarding the construct validity of the SDT measures when applied to interpret results of pain perception studies. Some of the criticisms and concerns about SDT analysis methodology and procedures, as applied to pain perception research, are discussed. These criticisms and concerns are categorised into three domains: Theory, methodology and interpretation. Possible solutions from recent research are outlined and these will inform the methodology of the studies within this thesis.

It is noted here that most of the sources used for this thesis' discussion on the 'SDT in pain research' debate were dated between the 1970s to the 1980s. This is reflective of the intense academic exchanges on the topic during that period. In fact, a recent exchange regarding the appropriateness of SDT methodology and interpretation for pain perception research indicates that the issues are still ongoing. (Clark, 2004; Gracely, Clauw, Ambrose & Petzke, 2004).

4.2 Theoretical concerns about SDT in pain research

SDT has had major influences on different areas of experimental psychology since the application of SDT to visual detection by Tanner & Swets (1954). However, pain perception has been one area that SDT has had little influence on its theoretical development. This chapter will examine some of the factors that may have prevented more widespread usage of SDT within pain perception research.

One of the factors was the interpretational ambiguity of change in pain perception outcome through the SDT measures. This factor will be discussed further in Sections

4.2.1 to 4.2.4. One of the main critics of the method in which SDT was applied to pain perception research was Professor Gary Rollman. Rollman's (1977) influential article criticised the theoretical assumptions and methodological problems when SDT was applied to pain research. The following sections will describe the main objections by Rollman and the other critics and commentators (Ominsky, 1979; Hayes, Bennett & Mayer, 1975; McBurney, 1975, 1976; Jones, 1979; Coppola & Gracely, 1983)

4.2.1 Criticism Number 1: Analgesia may be induced, but discriminability may remain unchanged.

Within research, certain assumptions on the conduct and interpretation of the results of the research are usually held by the researchers. This is independent of whether these assumptions were formulated consciously or unconsciously. Based on the SDT pain studies performed before Rollman (1977)'s critique, he asserted that SDT pain researchers held several unstated assumptions concerning the way SDT pain was studied. The assumptions were as follows (Rollman, 1977):

Assumption 1: A reduction in neural activity produced a reduction in experienced pain.
Assumption 2: A reduction in neural activity produced a reduction in the SDT discriminability index.
Assumption 3: A reduction in the SDT discriminability index indicated a reduction in experienced pain.
Assumption 4: A reduction in experienced pain will be reflected in a reduction in the SDT discriminability index.

These logical syllogisms will lead the reader to believe that SDT pain researchers did actually assume that a decrease in neural activity led to a decrease in experienced pain, and this in turn was manifested in a decreased discriminability index. Rollman (1977) used this logical preamble as the platform for building his objections to the way SDT was applied to pain perception. He did not object to SDT being used in pain perception research. His position was that d' only provides information about discrimination ability but little about pain perception.

Rollman (1977) then constructed a thought experiment to refute these assumptions. His thought experiment was as follows: Imagine an experiment examining a discrimination task during a pain modulation procedure. Assuming that the treatment's analgesic property in modulating the neural activity was not in doubt, based on assumption 2 posited by Rollman (1977), the analgesic will cause a reduction in the SDT discriminability measure. Returning to the SDT decision space referred to in Section 2.4.1, Figure 2.1 (reproduced in Figure 4.1), this can be represented by an increase in overlap of the Gaussian distributions which are internal representations of the physical stimuli (Figure 4.1). As the distance between the peaks of the distributions is reduced, a concurrent reduction in discriminability also occurred. However, Rollman (1977) argued that assumption 2 need not follow. The analgesic may act not only on the right-hand distribution of Figure 4.1 but also act on the left-hand distribution. Therefore, in theory if the analgesic exerted an equal effect on the perception of both stimuli, both the distributions representing the stimuli could move leftwards in the same magnitude. This situation will manifest as a constant discriminability being maintained despite a true analgesic effect taking place. This leaves the experimenter with an interpretational conundrum. A successful pain modulation treatment may be represented by either a reduction of discriminability at best or no change in discriminability at worst.

4.2.2 Response to the 'induced analgesia with no discriminability change' dilemma

Chapman (1977) responded to Rollman (1977)'s comments on the assumptions ascribed to SDT pain researchers (Chapman, 1977). Chapman (1977) denied that the assumptions, forwarded by Rollman (1977), were held as unstated beliefs by pain researchers using the SDT approach. He argued that to do so was to erroneously say that SDT pain researchers attributed a physiological correlate to d' in studying pain perception. Chapman (1977) stated that SDT was viewed by his research group as a probabilistic and decision-making model and should be treated as such when interpreting the results from their experiments. Pastore et al. (2003) also stated that SDT is a statistical model for evaluating the variable of interest and does not inherently characterise a physiologically independent process. Chapman (1977) attributed the fundamental disagreements between Rollman (1977) and himself to

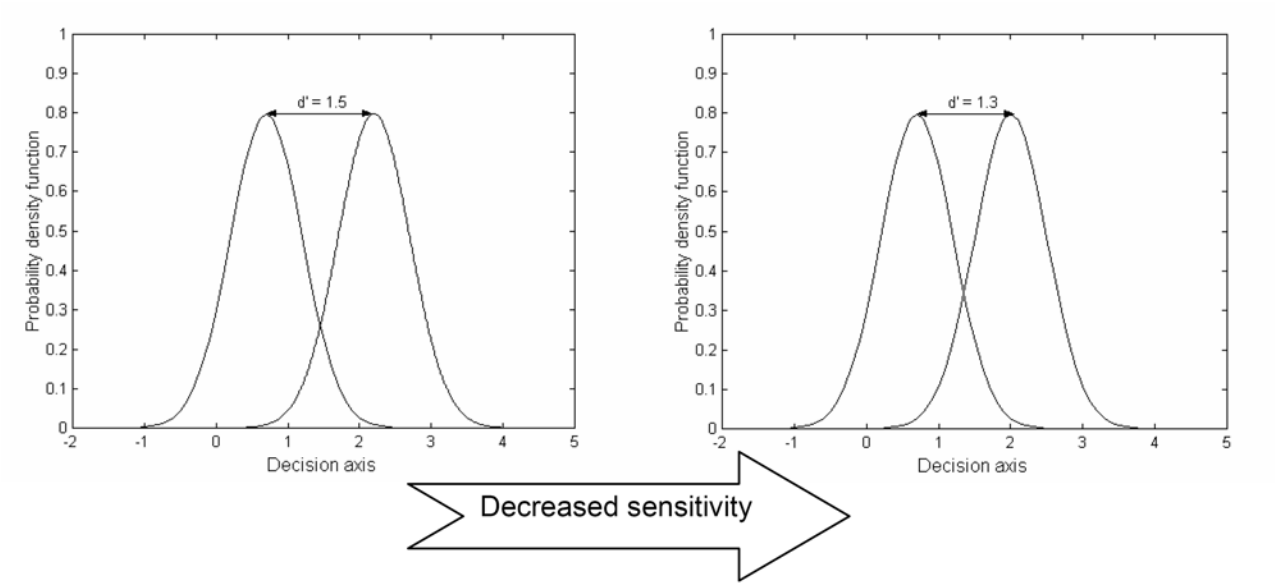

Figure 4.1. Figure A shows the internal representation of two stimuli. The graph represents the decision space of the participant in perceptually mapping the stimuli. The discriminability (*d'*) is the distance between the peaks of the normal distributions. Figure B shows that when the stimuli are perceptually harder to discriminate, this is internally represented by an increase in overlap of the distributions. The smaller distance between the peaks of the distributions signifies the lowered discriminability of the participant.

differences in perspective by which SDT was used to interpret empirical data. Chapman (1977) accused Rollman (1977) of attempting to caricature SDT pain research thereby making it open to numerous criticisms. Indeed, SDT pain researchers were often charged with the claim that sensory and psychological dimensions of pain perception can be separated through the use of SDT methodology. Fernandez & Turk (1992) reviewed the literature on SDT in pain research and concluded that discriminability and response bias were not pure measures of sensory and affective responses respectively. These two measures although statistically independent, may be functionally related (Fernandez & Turk, 1992; Macmillan & Creelman, 2005). This relationship was termed multicollinearity. It has been shown through computer simulation that discriminability may be affected by memory deficits and response perseveration (Coppola & Gracely, 1983). All of this implies that the conceptualisation of discriminability as a purely physiological function within pain research may not be entirely accurate under certain contexts. Therefore, descriptions of discriminability as a relatively pure physiological measure of pain and that

response bias measures all other non-sensory factors may be oversimplified statements associated with SDT measures.

In response to the imaginary situation of the maintenance of constant discriminability despite the presence of analgesia, Chapman (1977) deemed it extremely unusual. This was because no SDT theorist has considered this issue. Nevertheless, this research question has been approached in another way by Irwin & Whitehead (1991) using summated discriminability outcomes. Let us consider the discrimination task. The discriminability score obtained from the task represents the discrimination performance between two stimuli. Weber's law describes the relationship between the change in stimulus intensity that can just be discriminated by the participant (ΔT) and the starting intensity of the stimulus (T). The ratio between ΔT and T is a constant (k). Expressed formally,

$$\Delta T = k\,T \quad \text{or} \quad \Delta T/T = k\,.$$

The change in stimulus intensity that can just be discriminated by the participant is the same concept as the 'just-noticeable difference' (jnd) described in Chapter 1, Section 1.2. However, when the jnd between stimuli is used in SDT analysis, it is by definition the value of the physical stimulus corresponding to a d' of 1.0 (Irwin & Whitehead, 1991; Macmillan & Creelman, 2005). Assuming Weber's law is descriptive of pain perception processes, this implies that just-noticeable differences can be summated to obtain a continuum of the jnd describing the participant's discrimination ability along the stimulus range that is investigated. There is a precedent to this formulation in the form of the dol scale (Hardy et al, 1947, 1948). The dol scale consists of participants' reports of the smallest changes in intensity of the stimulus experienced. Each report of the smallest difference was taken as a jnd of the stimuli. Hardy et al (1947) found that the dol scale consisted of 21 such jnd. They defined one dol unit as two jnd reported by the participants. Therefore the dol scale consisted of 10 and one half dols. In a similar manner, the SDT discriminability of adjacent physical stimuli could be summated to obtain a representation of the total performance for the range of stimulus intensities presented to the participant (Braida & Durlach, 1972; MacMillian & Creelman, 2005). Using the analogy of physical distance, the range of stimuli represented internally by the participant is called the

'perceptual distance', and the index corresponding to summated discriminabilities is called the 'cumulated discriminability' (Braida & Durlach, 1972; Irwin & Whitehead, 1991; Macmillan & Creelman, 2005).

Irwin et al (1994) used the cumulated discriminability measure to show that perceptual distance may be a useful tool for indicating the presence of local anaesthesia. In that study, a topical local anaesthetic (Eutectic Mixture of Local Anaesthetics, EMLA®) was applied to the non-dominant forearm of participants. The four participants recruited were all authors of the paper. Electrocutaneous stimuli were administered to the participants via a constant current generator. The characteristics of the electrical stimulations were 200 ms duration, repetition rate of 500 Hz, pulse width of 0.2 ms. The five current amplitudes were separated by 1dB with a maximum of 414μA. This provided 4 pairs of currents for the SDT discrimination procedure. The participants underwent testing at the following periods: 5 minutes, 25 minutes, 45 minutes and 24 hours after the EMLA® was administered. The participants then rated their confidence on whether the lower or higher intensity of the stimuli was administered on a 6 point-scale. Discriminability for adjacent stimuli intensities within each time period were summated to obtain the cumulated discriminability. One interesting aspect of the experiment was the tracking of discriminability recovery from the topical anaesthetic effect. The idea behind this procedure was that if a depressed discriminability increased over time, this provided evidence that discriminability was valid as a construct to assess the recovery of cutaneous sensation.

The data supported the hypothesis that the discriminability of noxious electrocutaneous stimuli was reduced by the topical anaesthetic. Figure 4.2 shows the ROC functions of the pooled ratings of all the participants. The results also showed that the cumulative sensitivities increased for later periods of testing. This demonstrated that as the effects of the local anaesthetics wore off, the cumulative discriminability increased. This finding provided some evidence that discriminability may be seen as an indicator of analgesia. The reduction in cumulative discriminability also provided some evidence to disconfirm Rollman (1977)'s suggestion that discriminability may remain constant despite the presence of analgesia.

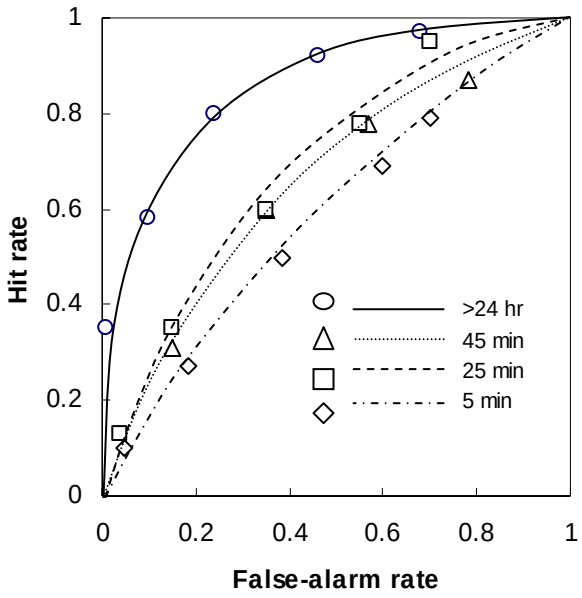

Figure 4.2 Receiver Operating Characteristic curves of the pooled discriminability of the participants in Irwin et al. (1994) experiment on the effects of topical local anaesthetics. The different symbols (○△□◇) show the ratings for different elapsed time after the removal of EMLA from the application site (adapted from Irwin et al. 1994)

One limitation of Irwin et al.'s (1994) study was that the research design did not contain a control condition. The recruitment of the authors as participants without the inclusion of a control condition meant that researcher expectation bias may have diminished the internal validity of the study. Nevertheless, one of the strengths of this study is the use of a topical local anaesthetic (EMLA®) with known and empirically tested analgesic properties for the induction of analgesia. This means that predictions about the effect of the topical local anaesthetic on changes of SDT measures can be made. This would establish the construct validity of SDT measures for demonstration of analgesia induced through topical local anaesthetics. Therefore, the study in Chapter 6 of this thesis used a topical local anaesthetic for the induction of analgesia in order to test if SDT measures reflect this change in sensory state. The difference between this study and Irwin et al's (1994) study is the inclusion of a within-subject control condition.

4.2.3 Criticism 2: Discriminability is reduced, but pain remains

Construct validity is related to the question of whether a particular test is measuring the variable of interest, or the extent that a measure reflects some underlying construct or latent variable (Hubley & Zumbo, 1996). The second criticism by Rollman (1977) was concerning the construct validity of noxious discrimination ability as a measure of pain perception. The argument was that SDT pain researchers obtained in their

research an index of discrimination between stimuli of different intensities, not pain. It is therefore theoretically possible that discrimination ability is worsened without pain being lessened. It is also possible that discrimination ability is unchanged yet pain is diminished. Rollman (1977) was therefore stating that the way researchers have defined the SDT measure of discriminability does not possess construct validity.

4.2.4 Response to the 'discriminability reduced but pain remains' argument

Rollman (1977) described discriminability in such terms that it appeared superficially not to provide the researcher with any information about pain perception. Rather, Rollman (1977) asserts that discriminability only provides information about discrimination. SDT essentially used a discrimination procedure which is not a common type of experimental task in current pain perception research. In other words, SDT lacked face validity for describing pain perception. Face validity is a concept describing the degree to which a measure superficially appears to be measuring a variable of interest (Domino, 2002). However, a measure lacking face validity does not immediately disqualify it as a potentially valid measure for assessing pain perception (Domino, 2002, p.57). It is perhaps more important that the construct validity of a measure be determined (Domino, 2002, pp.54-55).

One method of determining construct validity is through experimentation (Cronbach & Meehl, 1955). Logical inference and hypothesis formulation allows a researcher to make predictions on the possible outcomes for a particular variable of interest. In order to test the hypothesis, actual experimentation has to occur in order for the validity of the measure to be verified. Construct validation is a process and not simply a procedure. Therefore the process is ongoing and will be informed by future studies.

The context of the research may also influence construct validity of the SDT measures. For example, studies using SDT for evaluation of the effects of topical local anaesthetics demonstrated that discrimination ability possessed construct validity for assessing local anaesthesia to noxious electrocutaneous stimuli (Lineberry & Kulics, 1978; Irwin et al, 1994). Discrimination ability may also relate to some extent the influence of anxiety and attentional shifts in pain perception (Malow, 1981; Malow et al, 1989; Schumacher & Velden, 1984). These are two different contexts in which SDT measures were used. Discrimination ability in each context is associated

with slightly different meanings of what constitutes discrimination ability. It is theoretically possible that the SDT measures may possess construct validity in one context but not in another. This shows the importance of testing construct validity of SDT measures in the specific context of the research. Therefore, this thesis has examined the construct validity of the discrimination ability in two different contexts: 1) analgesia induced by a topical local anaesthetic compared to a control condition (Chapter 7) and, 2) chronic low back pain sufferers' discrimination ability to noxious thermal stimuli compared to healthy individuals (Chapter 8).

4.3 Methodological issues for SDT in clinical and applied research

The criticisms put forward by Rollman (1977) were not confined only to theoretical arguments. Concerns about the methodological soundness of the procedures employed by SDT pain researchers were also mentioned. Research using SDT has been adapted to suit the research question and the availability of time and resources in the applied setting. These problems will be considered together with the possible adaptations and modifications to the SDT methodology that would address them.

4.3.1 The number of trials within a discrimination task

Each presentation of the stimulus to the participant is called a 'trial'. SDT research has typically used large number of trials for testing. Green & Swets (1966) suggested that at least 250 trials per stimulus intensity were required, although they conceded that some sensory or perceptual problems may not be studied with such a large number of trials. Also, the participants in some domains of research may not be capable of performing more than a few trials. One such domain is the study of infant behaviour (Bargones & Werner, 1994). The practical limitations within these research areas may require some modification of trial presentation procedures. This section will discuss some of these practical limitations for pain perception research and the modifications required for SDT methodology.

SDT methodology in pain research has typically utilised a low number of trials. The number ranged from 8 trials per stimulus intensity (Yang et al., 1985) to 100 trials per stimulus intensity (Harkins & Chapman, 1977). This latter is still considered low in comparison to that suggested by Green & Swets (1966). The use of low numbers

within pain research may be due to several practical and ethical reasons. Firstly, the application of prolonged repetitive noxious stimuli to the participant may not be considered ethically acceptable unless the relatively large number is justified for the particular study (Charlton, 1995). Secondly, the application of a large number of trials is extremely time-consuming and most clinical research may be constrained by the participant's tolerance of the procedures involved. Lastly, hyperalgesia may become a concern after prolonged repeated stimulation (Møiniche, Dahl & Kehlet, 1993; Pedersen & Kehlet, 1998a, 1998b).

If a low number of trials was introduced within the research design of a study, two possible consequences may result. Firstly, if the rating design was used, there is a higher probability of some response categories not being utilised. Secondly, the SDT measures obtained from low trial numbers may deviate from the 'true' value as compared to measures obtained with a large number of trials. This deviation from the 'true' value of the measure is termed 'statistical bias'. It should be noted that statistical bias is different to the SDT measure of response bias. These consequences may not be trivial. It is worth considering their impact on the research design. If possible, solutions should be found to counteract the undesirable effects these consequences may have on the data analysis. The following section will consider the impacts and possible solutions of the identified problems.

4.3.2 Under-utilisation of the response set categories

Under circumstances of low trial numbers, it is possible that underutilisation of response set categories may occur. Consider the one-interval discrimination task for two stimulus intensities (see Section 2.7.3). An experimenter has determined that 20 trials per stimulus intensity will be used for the experiment. The response set has 6 categories indicating the confidence-ratings of the participant (Figure 4.3, upper figure). Assuming that the participant did not fully utilise all the categories available in the response set, this will mean that certain categories will have zero responses (Figure 4.3, lower figure). The construction of a receiver operating curve is achieved by converting the cumulated proportions of the categories for both stimulus intensities. If any of the categories are not used by the participant for the experiment, this may cause an underestimation of the ROC curve compared to the true ROC curve if all categories are used. This is also paralleled by an underestimation of the

discriminability. Fig 4.4 shows a graphical representation of this underestimation of discriminability.

1	2	3	4	5	6
I am absolutely certain the weak stimulus was presented	I am fairly certain the weak stimulus was presented	I am somewhat certain that the weak stimulus was presented	I am somewhat certain that the strong stimulus was presented	I am fairly certain the strong stimulus was presented	I am absolutely certain the strong stimulus was presented

Frequency of each response for each stimulus

	"1"	"2"	"3"	"4"	"5"	"6"
Higher intensity	0	0	10	10	0	0
Lower intensity	0	0	18	2	0	0

Figure 4.3. The upper figure shows a confidence-rating scale. The categories describe the subject's level of confidence in determining whether the stronger or weaker stimulus was presented in the trial. The lower figure shows the frequency of responses allocated by the participant to each category of the scale for a discrimination experiment.

There are two methods used to circumvent this problem. The first method is to correct the zero responses by giving the offending categories a nominally small number. Using this 'correction method', the number of categories remains unaltered. The two commonly used corrections are the '1/2*N*' and 'log-linear' corrections (Hautus, 1995; Macmillan & Creelman, 2005). The 1/2N correction is conventionally used in psychophysical research and the log-linear correction was suggested by Murdoch & Ogilvie (1968). The 1/2*N* correction simply replaces the zero proportions with the numerical value of 1 divided by 2 times the number of trials per intensity. In the one-interval discrimination task example, there were 20 trials per intensity. If a zero proportion exists, a correction of 1/(2 x 20) = 1/40 will replace the zero value of the category. In the case of the log-linear correction, a small numerical value will be added to all the categories to increase the total number of trials for that stimulus

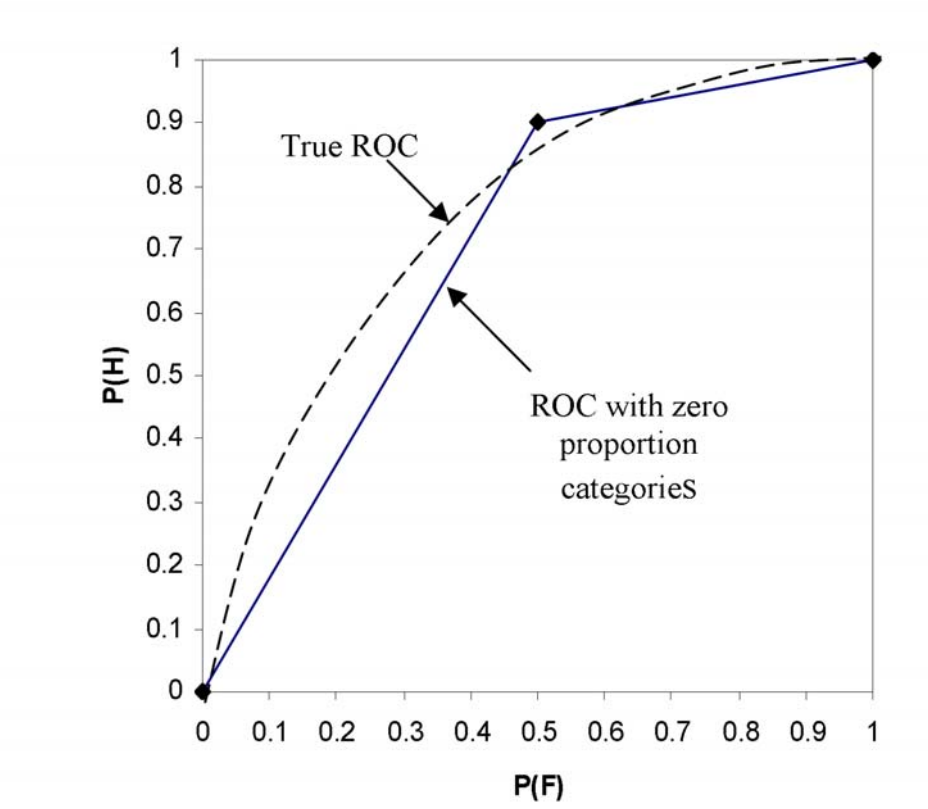

Figure 4.4. The figure shows the empirical ROC plotted from the response data in Figure 4.3. Note that only one data point is used to plot this ROC because of the presence of categories with zero responses. The theoretical ROC (true ROC) is superimposed on the empirical ROC to demonstrate that an underestimation of the actual area under the true ROC. The area under the ROC is indicative of the perceptual performance.

intensity by 1. In other words, the value $1/M$ (where M is the number of categories within the response set) is added to all the categories. Therefore, if the log-linear correction was used in the example (Figure 4.4), all the categories will have a value of 1/6 added to them. Although the problem of zero proportions has been resolved, the concern now is whether the corrections will distort the SDT measure in any way that may affect the conclusions of research findings. Hautus (1995) found that the corrections did cause the discriminability measure to either underestimate or overestimate the actual discriminability value. In other words, statistical bias was introduced through the use of the corrections. The bias was more severe when the actual discriminability value was either small (when $d' < 0.5$) or large (when $d' > 3.0$). Hautus (1997) also found that this bias was related to the number of participants used in the study and the number of trials per intensity used. With relatively smaller

numbers of participants and relatively smaller number of trials per intensity used, the statistical bias problem was worse. It was also evident that the log-linear correction produced less bias estimates of the discriminability than the 1/2*N* correction (Hautus, 1995).

The second method is to collapse the categories containing zero responses. This is a simpler procedure to perform and is recommended for correction of underutilisation of rating categories (McNicol, 1972; Hautus, 1995). For the illustrating example, the categories containing zero would be eliminated. The *d′* would be computed based on the remaining categories. When this procedure is performed, the number of response biases generated from the data in Fig 4.3 is reduced from 5 to 1. The advantage of this method is that statistical bias is less likely to be introduced to the SDT discriminability measure as compared to the log-linear and 1/2*N* correction methods (Hautus, 1995).

If the problem of zero proportions is persistent in the data and the use of corrections or category collapse is unavoidable, the best compromise is to preferentially use the correction or collapse procedure that yields the least amount of bias. In this case, the preferred method would be the category collapse method, as suggested by Hautus (1995) and Irwin & McCarthy (1998). Therefore, the studies in this thesis will use category collapse for correcting the presence of zero proportion categories.

The correction of zero proportions in response data raises another issue: the comparability of response biases generated from corrected data between participants, groups and studies. When the category collapse method is used, the total number of response biases will be reduced. The number of response biases is $(M-1)$ (where M is the number of categories within the response set). The number of categories left over from the collapse procedure is dependent on the number of zero proportions present in the participant data. Assuming that the number of collapsed categories is C, then the number of response biases generated from the leftover categories is $(M-C-1)$. Therefore the data set in Fig 4.3 will have a total of $(6-4-1) = 1$ response bias. It is likely that different participants will have different number of categories with zero proportions. Therefore, it is impossible to compare the response

bias between participants with different response biases or obtain a summary statistic for the response bias within a group. This issue is not resolved with the use of the correction methods of $1/M$ and $1/2N$. Although these correction methods retain the original number of $(M-1)$ response biases, the categories that are 'corrected' provide a response bias artefact. The presence of response bias artefacts means that a comparison of between participant or group response biases may produce misleading results. The use of large category numbers in the response set and small number of stimulus trials, the likelihood of obtaining zero proportions is high. However, there is no mention of correction methods used in previous SDT pain perception research. The problem relating to non-comparability of response bias results is also not mentioned in previous pain perception research using SDT methodology. This casts some doubts on the analysis of response bias in terms of the accuracy of comparisons for previous SDT pain studies. In view of the problems associated with the comparison of response bias when zero proportion categories are present, response bias was excluded as an outcome measure for analysis within the studies of this thesis.

4.3.3 Too many categories in the response set

A problem related indirectly to the zero proportion categories problem is the optimum number of categories used in the response set. Rollman suggested that SDT pain studies tend to utilise too many categories in the response set coupled with too few trials. For example, Yang et al (1991) used a total of 14 categories in their response set with only 8 trials per intensity used to test the participants. However, in recent research, the number of categories has been reduced to about 8 categories (Janal et al, 1994; Pertovaara, 2004; Soetanto et al, 2004).

McNicol (2005) stated that there were no rules in the choice of the optimum number of categories. However, several factors must be considered. Firstly, although a larger number of categories will provide a larger number of points for plotting the ROC curve, there is the chance that participants will not use the categories consistently. For example, Clark & Mehl (1973) found that participants, in their study, were not able to internally visualise the large number of criteria for the rating procedures and use them for producing judgments. This resulted in a lower discriminability compared with the predicted value. An approximate number of categories that a participant may use has

been suggested to be 7 plus or minus 2 categories (Miller, 1956). This number is often used as a rule-of-thumb estimate for the limitations on information processing by humans. Laming (1984, 1997) has also suggested that participants are able to reliably provide judgements for about 5 categories. Secondly, if a larger number of categories exist within the response set, it is more likely that some categories will not be used. This is a practical problem related to the experimental design, but one that can be easily resolved by either increasing the number of trials administered or using one of the correction procedures described in Section 4.3.2. It is suggested that a lower number of categories be used for the rating experiment. The category numbers may range between 5-9 categories. If larger numbers of categories are necessary, a proportional increase in trials administered should be preferably used in the study. The studies in this thesis used 6 categories for the response set.

4.3.4 Effect of stimulus-range on the participant's performance

The most common type of response set used in SDT pain perception studies requires the participants to provide ratings of perceived magnitude (Figure 4.5a). The ratings of perceived magnitude procedure will hence be referred to as the 'magnitude-rating' procedure or task. The other type of response set used is the 'confidence-rating' scale. The associated 'confidence-rating' procedure requires the participant to judge whether the higher or lower intensity stimulus was presented during the trial and the confidence used to provide this judgement (Figure 4.5b). The difference between these two types of tasks is apparent and it is expected that they also differ in terms of the discriminability obtained from SDT analysis. Interestingly, Clark & Mehl (1973) found that there was no significant difference between the discriminability from the magnitude-rating task and the confidence-rating task. Even Rollman (1977) agreed that the magnitude-rating procedure was an acceptable deviation from the usual types of discrimination task. In fact, magnitude-rating tasks can be classified under a special form of classification procedures. Macmillan & Creelman (2005) defined classification procedures as having participants "use M responses to sort N stimuli into categories". If there are 2 stimuli and 2 types of responses ($N = M = 2$), this is the 'yes-no' procedure discussed in Chapter 2, Section 2.3. If there are more possible stimuli than responses ($N > M$), this is traditionally called 'category scaling'. If the number of stimuli is the same as the number of responses ($N = M$), but the number of

stimuli and responses are greater than two, this is called ‘absolute judgement’, ‘absolute identification’, or simply the ‘identification procedure’.

A

Warm	Hot	Faint Pain	Painful	Very Painful	Severe Pain

B

1	2	3	4	5	6
I am absolutely certain the weak stimulus was presented	I am fairly certain the weak stimulus was presented	I am somewhat certain that the weak stimulus was presented	I am somewhat certain that the strong stimulus was presented	I am fairly certain the strong stimulus was presented	I am absolutely certain the strong stimulus was presented

Figure 4.5. Two forms of rating scale. A. A magnitude-rating scale. The subject is instructed to rate the perceived magnitude and qualitative description of the stimulus. B. A confidence-rating scale. The participant is instructed to rate his/her confidence in determining whether the stronger or weaker stimulus was presented during the trial.

If the number of response categories is more than the number of stimuli (N<M), this is called magnitude estimation. SDT pain studies usually utilise more types of responses than stimuli, therefore magnitude estimation, as defined by Macmillan & Creelman (2005), is used. The magnitude estimation procedure requires the participant to provide a number estimation of the stimuli. The magnitude-rating procedure could also be viewed as an ordinal representation of the magnitude estimation procedure (Braida & Durlach, 1972; Laming, 1984).

Theoretically, there are reasons to believe that the discriminability obtained from magnitude-rating may be lower than confidence-rating procedures. Assuming that the difference in stimulus intensity between any two adjacent stimuli are equal for all stimuli presented, it is known that when more than two stimuli are presented to the participant for judgement within one discrimination task, the participant’s attentional and memory capacity may be divided in an attempt to judge the larger range of stimulus intensities (Parducci, 1965). It may also be that the participants tend to make

their subjective ratings based on the context in which the stimulus was presented. Therefore, a stimulus of a certain intensity will be easier to judge in the context of a smaller stimulus range compared to a larger stimulus range (Pyn, Braida & Durlach, 1972; Poulton, 1989). SDT pain studies using magnitude-rating procedures usually present more than 2 stimuli, hence a larger stimulus range, for the participant's judgement within one discrimination task. For example, Yang et al. (1991) presented a stimulus range of 0-390 mcal/cm^2/s in the form of 4 stimulus intensities (0, 100, 340 and 390 mcal/cm^2/s). Although this procedure is efficient, the results showed that the theoretically predicted discriminability may be poorer than if a smaller stimulus range was presented. Irwin & Whitehead (1991) compared three one-interval psychophysical tasks involving noxious electrocutaneous stimuli. The tasks were a discrimination task judging 5 pairs of stimuli, an identification task (N = M) judging 6 stimuli, and a magnitude-rating task judging 6 stimuli. For the discrimination task, the stimulus range was 0.5 dB for each stimuli pair and the stimulus range for the identification and magnitude-rating tasks were 3.0 dB for all the stimuli combined. The results showed that discriminability was better for the discrimination task compared to the other tasks due to the stimulus range effect described above. Rollman (1983) has also shown that when four randomised stimuli per block of trials were compared to two randomised stimuli, the four stimuli set tended to yield a lower discriminability. Although magnitude-rating is a legitimate extension of the SDT judgement task, it may still be influenced by the effects of attention, memory as well as the effects of the context. If the researcher aims to minimise the effects of attention, memory and context on the discriminability of participants, the one-interval discrimination task judging only 2 stimuli at any one time may be the more appropriate choice.

4.4 Definition and interpretation of SDT measures

In Chapter 3, one of the problems faced by SDT pain researchers was the definitional problem of 'analgesia' within the framework of SDT. Critics of SDT pain research sought for clarification on what constituted analgesia when SDT measures were used. If analgesia was induced, was this indicated by a reduction in discriminability alone, an increase in response bias alone, or perhaps even a change in both the discriminability and response bias? This was a challenging question that SDT pain

researchers did not fully answer. Two reasons are proposed in this thesis for the apparent ambiguity to the analgesia question.

4.4.1 Misrepresentation of SDT pain researchers' views

The SDT pain researchers initiated their work based on the theoretical framework of the original SDT theorists (Green & Swets, 1966). They have modified the practicalities of experimental procedure to accommodate the requirements of pain research. For example, instead of the conventional confidence-rating response set, the magnitude-rating response set was used. Another modification was the reduction of the number of trials per intensity administered. Such improvisations were pragmatic. However, the theoretical framework was still essentially similar to the original SDT theory. The original SDT framework provided only a structural model for researchers to work on their hypotheses in their individual disciplines. It did not impose a rigid definition of what discriminability and response bias may constitute (Jones, 1979; Pastore et al., 2003). Therefore, discriminability might indicate neural functioning or it could indicate cognitive functioning, depending on the area of study. The definitions of the SDT measures were dependent on the researcher's theoretical persuasion, coupled with the design of the experimental set-up. It is perhaps helpful to review some of the definitions offered by the main proponents of SDT pain research. Direct quotations are drawn from the writings of Clark and Chapman to avoid misinterpretation of their original meanings.

Clark (1994, p.43) defined the SDT measures as the following: "SDT yields two measures of perceptual performance. The discriminability measures…reflect the accuracy with which a person …judges whether event A or event B has occurred. The report criterion measure…quantifies the subject's response bias, that is, the general tendency to report one of the events as occurring more frequently than the other" and "SDT provides…two separate measures: d', which is related to sensory discrimination, and the other, the report criterion, Lx, which is influenced by affective and other psychological variables" (Clark, 1994, p.56).

Chapman's (1976, p.269) definition of the SDT measures was slightly indirect. He assumed "… that decreases in d' for a subject perceiving normally painful stimulation in a properly structured experiment are indicative of a loss of pain sensibility, hence

they reflect analgesia…response bias, helps shed some light on the motivational aspects of the pain experienced in the laboratory context".

It is appropriate to mention that these definitions were offered in the context of pain modulation studies. Clark and Chapman both agreed on the definition that discriminability may indicate a sensory component, and the response bias may indicate a non-sensory (albeit slightly ambiguous) component of pain perception. Nevertheless, this came with the disclaimer that proper design of the experiment as well as intelligent discretion in interpretation of the results by the researcher was also needed. This suggested that Clark and Chapman did not interpret every discriminability decrease to indicate analgesia, or every response bias change to indicate some significant psychological influence. In this situation, Chapman (1977) was, in some sense, justified in saying that Rollman had misrepresented the positions of the SDT pain researchers with regards to the interpretation of discriminability. In an editorial for the Journal of Anesthesiology, Ominsky (1979) reiterated the position of Rollman that SDT pain researchers tend to infer that a decreased discriminability or discrimination ability was the same as analgesia. Clark & Yang (1980) strongly protested against this interpretation of their position. Clark & Yang (1980) then stated unequivocally that they did not interpret decreased discriminability to be analgesia, but suggested there was mounting evidence to show that it was a useful correlate. Perhaps readers and critics of SDT pain research had misunderstood the positions of the SDT pain researchers. This misunderstanding was, in some ways, also compounded by the complexity of SDT to researchers and clinicians unfamiliar with the theoretical underpinnings of SDT. This led the readers and critics to interpret the research findings based on their conception of what SDT pain research aimed to achieve.

4.4.2 Interpretational ambiguity of magnitude-rating scale

In Section 4.3.4, it was mentioned that the large stimulus-range used in magnitude-rating procedures may result in poorer performance compared to discrimination tasks that utilised a smaller stimulus-range. Although not mentioned by any of the critics, it may be possible that the use of the magnitude-rating scale might have also contributed to the interpretational ambiguity of the changes in SDT measures. Coppola & Gracely (1983) stated that it was the interpretation of changes in SDT measures that caused

problems, but not the interpretation of no change. This thesis suggests that part of the interpretational problem of change may be attributed to the use of the particular type of scale.

Before this section discusses how the magnitude-rating scale might have contributed to the ambiguity question, first consider the practical outcomes that might potentially be encountered when using SDT indices for measuring discrimination under the influence of a putative analgesic. Figure 4.5B shows a confidence-rating scale. When the confidence-rating scale is used, two likely outcomes may be produced, assuming the analgesic effect is unidirectional: The first outcome is that there is a lowered d' with either a lowered, similar or raised response bias between the intervention and control groups. This outcome could be interpreted to be a successful induction of analgesia within the intervention group. The response bias shifts may be interpreted to be the groups' unique usage of the confidence-rating scale. Therefore the response bias shift is interesting in its own right if certain aspect of judgment behaviour is the topic of interest. However, for this particular example, the response bias is interpreted as the participant's level of confidence in deciding on the choice of one of the categories. Therefore, the assumption is that response bias changes do not affect the interpretation of the d' findings. The second possible outcome would be that the d' remained the same with a shift in response bias. This would suggest that the analgesia induction was unsuccessful. The response bias shift in this example is also of no consequence to the interpretation of the d' outcome. Some researchers have adopted this approach when investigating the effects of analgesics (Irwin & Whitehead, 1991; Irwin et al., 1994)

Signal detection theory indices outcomes	
Discrimination ability (d′)	Response bias (c)
Lowered	Raised
Lowered	Similar
Lowered	Lowered
Similar	Raised
Similar	Similar
Similar	Lowered
Raised	Raised
Raised	Similar
Raised	Lowered

Table 4.1. All theoretically possible combination of outcomes for discrimination ability and response bias changes after an intervention..

When the magnitude-rating scale is used, nine possible combinations of discriminability and response bias changes may be possible. Figure 9.1 shows the matrix of these discriminability and response bias combinations. However, to demonstrate the more ambiguous interpretation when using this scale, only one combination will be considered with two possible interpretations offered. The combination considered is when similar *d′* were obtained for both the intervention and control groups, with a relatively higher response bias for the intervention group. The first interpretation of this result would simply be that no analgesia was induced based on the similar *d′* obtained. The change in response bias may be considered to be irrelevant to the interpretation of the *d′* outcome. However, due to the descriptors used in the magnitude-rating scale, a higher response bias would mean that the participants have a tendency to report a less intense experience. This interpretation of the relevance of response bias has seldom been adopted by previous pain studies using SDT. Most studies either took the response bias change to indicate some kind of perceptual change or interpreted it to be linked to some aspects of cognition or affect (Clark, 1974; Chapman & Butler, 1978; Clark et al., 1981). Therefore, it is unlikely that any investigators would ignore the response bias changes entirely when using the magnitude-rating scale.

More controversially, it could also be concluded that the raised response bias indicated that analgesia was induced, even though this was not reflected in a *d′*

change. This second interpretation of the result takes into account the change in response bias. The argument is that a raised response bias would indicate approximately that more responses were made by the participants in the categories with less painful descriptions. Therefore, this could mean there has been some form of analgesia so that the noxious stimuli were rated less painful. This interpretation has been adopted by several studies (Yang et al., 1979; Lineberry & Kulics, 1978). When interpretations for all nine combinations of discriminability and response bias are considered, the analysis becomes rather unwieldy and complicated.

With the use of a magnitude-rating scale, the interpretation of the results therefore becomes problematic from the possible functional dependence between the discriminability and response bias measures (Fernandez & Turk, 1992). Therefore, the magnitude-rating scale and confidence-rating scale may not be equivalent, since the interpretation of both scales' likely outcomes appears to differ in complexity. This position has been tested empirically by Rollman (1983). In his study, Rollman (1983) administered noxious electrocutaneous stimuli to the participants and then they were asked to make judgements based on four different types of category scales consisting of (Figure 4.6): (1) a pain intensity scale, (2) a stimulus intensity scale, (3) an unpleasantness scale, and (4) a discrimination scale. The results showed that the discrimination scale, which is similar to the confidence-rating scale discussed in this section, always yielded a higher performance for the different SDT discriminability indices calculated. Rollman (1983) did not suggest the mechanism for this result but it did demonstrate the need for experimenters to evaluate the choice of the rating task employed in experiments. The d' differences for the various scales point to a further difficulty when comparing studies that use different scales. In order to avoid potential interpretational ambiguity, the confidence-rating scale was used for the studies in this thesis instead of the magnitude-rating scale. Nevertheless, it was important that the outcomes generated by the confidence-rating scale could be compared to the outcomes obtained by previous studies, which had mostly used the magnitude-rating scale. A theoretical link was required to link the two approaches in order that some form of comparison can be made between the two scales. Chapter 6 outlines an analytical framework suggested by Braida & Durlach (1972), Irwin & Whitehead (1991) and Laming (1984) for analysing the different scales and offer a common framework for comparing the results.

	Dimension			
Rating	Pain	Intensity	Unpleasantness	Discrimination
1	Nothing	Nothing	Nothing	Certain it was the weaker
2	Touch	Possible sensation	Not unpleasant	Fairly sure it was the weaker
3	Very faint pain	Very fainy intensity	Very faint unpleasantness	Unsure
4	Faint pain	Faint intensity	Faint unpleassantness	Fairly sure it was the stronger
5	Mild pain	Mild intensity	Mild unpleasantness	Certain it was the stronger
6	Moderate pain	Moderate intensity	Moderate unpleasantness	-
7	Strong pain	Strong intensity	Strong unpleasantness	-
8	Very strong pain	Very strong intensity	Very strong unpleasantness	-

Figure 4.6. The category rating scales used for Rollman's (1983) study. It was found that the 'Discrimination' scale generally yielded higher discriminability compared to the other scales.

4.5 Critique of past SDT studies

Although Rollman's (1977) paper was critical of the methodology and interpretation of discriminability used within SDT pain research, his paper has also raised pertinent issues. However, an overview of the SDT pain research published after Rollman's critique (Chapters 3 to 4) showed that many of the issues raised were not directly addressed empirically.

For example, some of the methodological and analytical issues raised by Rollman include the large number of response set categories used, the small number of stimulus trials. This thesis has stated that these two issues would also have an influence on the statistical bias introduced into the SDT discriminability index chosen for the study. Clark (1994) responded to the number of stimulus trial issue. He stated that if the trial number is sensitive enough to allow a computation of a statistically significantly discriminability index when a statistical comparison is made, then the trial number is sufficient. Clark's (1994) response is problematic. Firstly, the statement does not account for the issue of statistical bias introduced into the

discriminability index. If the purpose of the study is to provide an accurate description of the participants' discrimination ability, a biased discriminability index will mislead the researcher in the interpretation of the study results. Secondly, statistical bias may also be introduced when the likelihood of zero response categories increases. Clark (1994) cites the example of the study by Yang et al. (1985) to support his statement. Yang et al's (1985) study is also described in Section 3.2, Chapter 3 of this thesis. Clark (1994) mentioned that 8 stimulus trials per stimulus intensity, for 4 stimulus intensities, was sufficient to elucidate the statistically significant differences in discriminability between chronic pain patients and healthy participants for Yang et al's (1994) study. However, 14 response categories were used for this cited study. This meant that the response data for each of the stimulus intensities would have contained at least 4 zero response categories. Yang et al's (1985) did not describe the procedure taken to correct the zero responses. Therefore, the extent of statistical bias introduced into the discriminability due to the presence of zero response categories in the data is unknown. The issue of statistical bias within the discriminability obtained is not trivial. If the accuracy of the data is questioned, then it would also follow that any analytical results obtained from biased data should also be questioned.

4.6 Summary

The above sections have discussed three areas of concern for SDT pain research: theoretical, methodological and interpretational. The theoretical problems faced were the construct validity of discriminability as a measure of pain perception. This issue was deemed to be solvable only through careful experimentation. Several methodological problems were discussed and these concerned the number of trials within the experiment, underutilisation of response set categories, too many categories used for the response sets and the effect of stimulus-range on the participant's performance. These concerns are not trivial but they are also easily solved with the recommended solutions of correction for zero proportions, reducing the number of categories and the comparison of only two stimuli within each experimental block. The interpretational problems were attributed mainly to the misrepresentation of SDT pain researchers' theoretical positions, as well as the use of the magnitude-rating scale.

4.7 Conclusion

This chapter has expounded on the problems and concerns faced by SDT pain research. Most of the problems were not as intractable as they seem, however, neither were they satisfactorily answered by SDT pain researchers. The debate may seem dated from the dates of the articles drawn for this review. However, recent correspondence and papers suggested that this debate has reached a stalemate and no further progress made in addressing the problems (Craig & Rollman, 1999; Clark, 1994; Clark, 2004; Gracely et al, 2004). A positive viewing of the criticisms is that it has laid the foundation for future investigations on SDT in pain research. In order for SDT pain research to move forward, an adequate rejoinder to the above problems needs to be provided by a new generation of researchers.

Chapter 5
Methodological Studies

5.1 Introduction

Most of the studies reviewed in Chapter 3 and 4 using SDT methodology involved the use of experimental noxious stimuli. There are different methods of administering experimentally noxious stimuli. Gracely (2006) described five physical methods using heat, cold, ischaemia, mechanical and chemical stimuli. The studies for this thesis used heat for the induction of noxious stimuli. The heat stimuli were administered via the Somedic Thermotest (Somedic A.B., Sweden), a quantitative sensory testing machine. This chapter will be establishing the accuracy and precision of the temperatures produced by the Somedic Thermotest before and during the conduct of the studies for this thesis.

The aim of measurement is to know the "true value" of the quantity of interest. However, no measurement is perfectly accurate or exact. The deviation from the "true" values of the measured quantities can be due to multiple factors. One of these factors is the uncertainty inherent in the measurement itself. Another factor is the statistical bias that may be introduced through the number of data points collected throughout the experiment. The term 'uncertainty' is used in a statistical manner which describes dispersion of quantitative values that are attributed to a quantity of interest (International Organisation for Standardization, 2004). The establishment of the amount of uncertainty in the measurements of a specific quantity from test equipment and procedure are important for confidence in the measurements acquired. It is therefore crucial that the characteristics of the uncertainty within measurements are specified explicitly through experimentation.

5.2 Definitions

The International Organisation for Standardisation (whose recommended acronym is ISO) and the National Institute of Standards and Technology (NIST) have both published documents providing the terminology and guidelines for measurement in scientific fields. Some of the more common terminology will be outlined here.

When a variable is quantified through experimentation, the variable will be expressed in the unit of measurement appropriate to the variable. The specific “quantity intended to be measured” is called the “measurand” by the International Vocabulary of Basic and General Terms in Metrology (ISO, 2004). For example, the determination of the boiling point of water requires the quantification of temperature. Thus temperature is the measurand in this instance. And the units of measurement are Kelvin (K) or degrees Celsius (°C) The term “measurand” is the basis of other definitions pertaining to measurement. The measurand of the present accuracy and precision study is also temperature and the unit of measurement is degrees Celsius (°C), the conventional unit used in thermal quantitative sensory testing studies (Lindblom, 2005).

Central to the principles of measurement are the concepts of “accuracy” and “precision”. These concepts provide the foundation for acceptance of measured quantities within an experiment, at the local experimental context, and acceptance of theories and models, at the philosophical context. Therefore it is crucial for the accuracy and precision of the measurements to be determined, or at least estimated.

Accuracy or accuracy of measurement is defined as the “closeness of agreement between a quantity value obtained by measurement and the true value of the measurand” (ISO, 2004). Accuracy is therefore the nearness of the measured values with the reference or “true value” of the measurand. A measured value that is hugely different from a reference value is termed to be low in accuracy. A measured value that is near to a reference value is known to be high in accuracy. The accuracy of a measurement is a relative concept. Only by comparison to another standard or reference can the accuracy of the measured value be defined. In this study, the reference will be the absolute desired temperature to be calibrated. For example, if the averaged measured temperature of the equipment thermode surface is 47.01°C ± 0.01°C, for the reference of 47 °C. The “closeness of agreement” between the measured and reference temperatures, and therefore the accuracy, is considered good.

Another concept that accompanies, but is independent of accuracy, is precision. Precision or measurement precision is the “closeness of agreement between quantity values obtained by replicate measurements of a quantity, under specified conditions” (ISO, 2004). Precision is the nearness of the repeated measurements with each other.

Therefore, precision is related to the concept of reliability where repeated measurements close in values are said to be high in precision and measurements remote in values are termed low in precision.

In the design of experiments to determine precision of the measurements, the pre-defined conditions of experimentation are important for determining the types of precision of measurement. Generally, there are two types of precision: the repeatability and reproducibility of measurements. The repeatability condition is defined as the 'condition of measurement in a set of conditions including the same measurement procedure, same operator, same measuring system, same operating conditions and same location, and replicated measurements over a short period of time' (ISO, 2004). The reproducibility condition is defined as the 'condition of measurement in a set of conditions including different locations, operators, and measuring systems' (ISO, 2004). Therefore, measurement precision under the repeatability and reproducibility conditions are slightly different to each other based on the factors that are kept constant or varied.

Included within the definition of repeatability is the variable of time between the repeated measurements. It is entirely possible that a slow change may be manifested in the measurements due to either time or variability within the instruments. This phenomenon is known as drift (ISO, 2004). If the drift is associated with time, it is termed temporal drift. The process of drift as a form of uncertainty cannot be eliminated. One of the solutions is to monitor the effect of drift on the measurements and correct the effect through calibration of the equipment.

Experiments A, B and C, described in this chapter, were conducted to estimate the accuracy and precision of the temperatures produced by the Somedic Thermotest (Figure 5.1A), the equipment used for all the studies within this thesis. For measurement precision, both the repeatability and reproducibility conditions were examined through three experiments. The extent and nature of the temporal drift between two adjacent experiments was also estimated.

Before the experimental procedures are described, the use of Quantitative Sensory Testing in pain research will be briefly outlined.

5.3 Quantitative sensory testing in pain research

The Somedic Thermotest is a Quantitative Sensory Testing (QST) machine. QST machines have been increasingly used for characterising sensation loss, experimental pain threshold, allodynia and hyperalgesia (Lindblom, 2005). The Thermotest uses thermal stimuli as the modality of testing the subject's cutaneous sensation. The thermal stimuli are produced through the fluid-cooled Peltier element within the thermode. This mechanism of cooling and heating allows a fairly rapid rise or drop in temperature. This effect is especially useful in adaptive psychophysical procedures such as the Method of Limits to estimate the subject's threshold for heat or pain. The Method of Limits procedure consists of the heat stimuli increasing or decreasing in temperature administered on a selected region of the participant's body. The participant is required to indicate when the thermal stimulus becomes just detectably painful (heat detection threshold) and when the stimulus becomes 'faint pain' (pain detection threshold). The heat and pain detection thresholds are usually reported in the physical units of measurement, in this case °C.

Although the Method of Limits is commonly used in pain perception studies, the rating experiment procedures utilised within this thesis required several fixed temperatures to be maintained. The participants were required to discriminate between the different static temperatures and make verbal judgments on which temperature was administered for that particular trial. Temperature settings on QST machines allow static temperatures above pain threshold, also called suprathreshold stimuli, below pain threshold to be pre-programmed. The machine's ability to sustain the fixed temperatures is valuable in psychophysical procedures, albeit underutilised in pain perception studies. The fixed temperature serves as an objective standard by which the participant's subjective responses may be compared. Suprathreshold thermal stimuli can be used to quantify the nociceptive responses of the individuals. This requires the use of temperatures that are intense enough for the stimulation of the cutaneous nociceptive afferent fibres. The Somedic Thermotest has a temperature production range of 5°C to 56°C. Psychophysical studies have found that the temperatures required for activation of nociceptive afferent fibres range between 40°C to 49°C (Lamotte & Campbell, 1978; Meyer & Campbell, 1981; Van Hees & Gybel,

1981; Yarnitsky, Sprecher, Zaslansky & Hemli, 1995). Therefore, the Thermotest is capable of producing temperatures that activate nociceptive responses in participants. The production of fixed temperature is a novel use of the QST machine. Due to the novelty of the procedure, information on the accuracy and precision of the Somedic Thermotest were not available from the user's manual of the equipment (Somedic A.B., 2002). Experiments A, B and C, within this chapter, examined the accuracy and precision of the temperatures produced by the Somedic Thermotest.

5.3.1 High versus low trial numbers in sensory testing

Rating experiments utilising discrimination procedures present stimuli repeatedly to determine the perceptual discrimination ability of subjects. In a laboratory setting, high stimulus presentation numbers, also called high trial numbers, are usually utilised. However, such an endeavour may not be practicable in non-laboratory settings.

Several factors may mitigate the use of high trial numbers in discrimination procedures using noxious stimuli. Firstly, hyperalgesia may develop due to repeated thermal stimulation (Pedersen & Kehlet, 1998a, 1998b). The onset of hyperalgesia may distort the 'true' value of a measure representing the subject's noxious discrimination ability. Secondly, lower trial numbers may lower the statistical precision and accuracy of the result (Hautus, 1997; Millers, 1996). Thirdly, when rating designs are used for the discrimination procedures, lower trial numbers increase the likelihood of yielding zero response categories within the rating response set. In other words, the subject may either assign all or none of his/her responses to one category. However, there is little evidence regarding the possible biasing effect of low stimulus numbers on noxious discrimination ability measure. Experiment D, within this chapter, compared two trial numbers to determine the effect of trial number on the discrimination ability measure. The relative efficiency, an index of statistical precision, of the procedures was computed to describe the statistical variability of the discrimination ability between the two stimulus numbers. A statistical comparison, between the variability of discrimination ability for the two trial numbers, was also performed to determine if the variability difference is statistically significant.

5.4 Aims of accuracy and precision experiments

A total of four experiments were therefore conducted to: a) examine the accuracy and precision of the temperatures produced by the Thermotest and, b) compare the discrimination ability variability between two trial numbers (17 trials and 40 trials per stimulus intensity). The aims of the four experiments are as follows:

- Experiment A:
    - to estimate the accuracy and precision of the temperatures produced by the Thermotest for the pre-calibrated and post-calibrated states.
- Experiment B:
    - to estimate the accuracy and precision of the temperatures produced by the Thermotest for the pre-calibrated and post-calibrated state. Experiment B was conducted 6 months after the conduct of Experiment A.
    - to determine the repeatability of the temperature measurements between the post-calibrated state of Experiment A and pre-calibrated state of Experiment B.
- Experiment C
    - to estimate the accuracy and precision of the temperatures produced by the Thermotest for the pre-calibrated and post-calibrated state. Experiment C was conducted 9 months later from the conduct of Experiment B. Experiment C was also conducted at a different location (Western General Hospital Edinburgh) to Experiments A and B (Queen Margaret University, Leith campus)
    - to determine the reproducibility of the temperature measurements between the post-calibrated state of Experiments B and pre-calibrated state of Experiment C.
    - To determine the reproducibility of the temperature measurements between the test site of Experiments A and B to Experiment C
- Experiment D
    - To determine the relative efficiency of the procedure using a high trial number (n = 40) compared to the procedure using a low trial number (n = 17)

- To examine if the variabilities of the discrimination ability measure for the two trial numbers are statistically significantly different.

Experiments A, B and D were conducted at Queen Margaret University, Leith campus, the test site for two of the studies within this thesis (Chapters 6 and 7). Experiment C was conducted at the Western General Hospital Edinburgh, the test site for one of the studies within this thesis (Chapter 8).

Experiment A

5.5 Aim

a. To estimate the accuracy and precision of the temperatures produced by the Thermotest for the pre-calibrated and post-calibrated states.

5.6 Methods

5.6.1 Experiment environment

The experiment was conducted in an enclosed rectangular room within Queen Margaret University. The room was not noise- or temperature-controlled. However, the noise interference was minimal. Any variations in the room temperature were monitored by temperature measurements taken immediately before and after the experiment. A standard mercury thermometer was used to measure the room temperature. The room had a volume of 35.90m^3.

5.6.2 Thermotest and calibration equipment set-up

The calibration equipment was designed by the manufacturers of the Thermotest. The calibration equipment consists of an aluminium frame, a thermocouple, a thermocouple probe and the calibration software, EXPOSURE (Somedic A.B., Sweden). The calibration equipment is shown in Figure 5.1A. The calibration software, installed on a Toshiba Satellite Pro laptop, was used in all experiments described within this chapter. The laptop had a 15 inch TFT ISD screen, Intel Celeron 1.2GHz processor and 128Mb RAM. The Testo 950 thermocouple with a flat-tipped

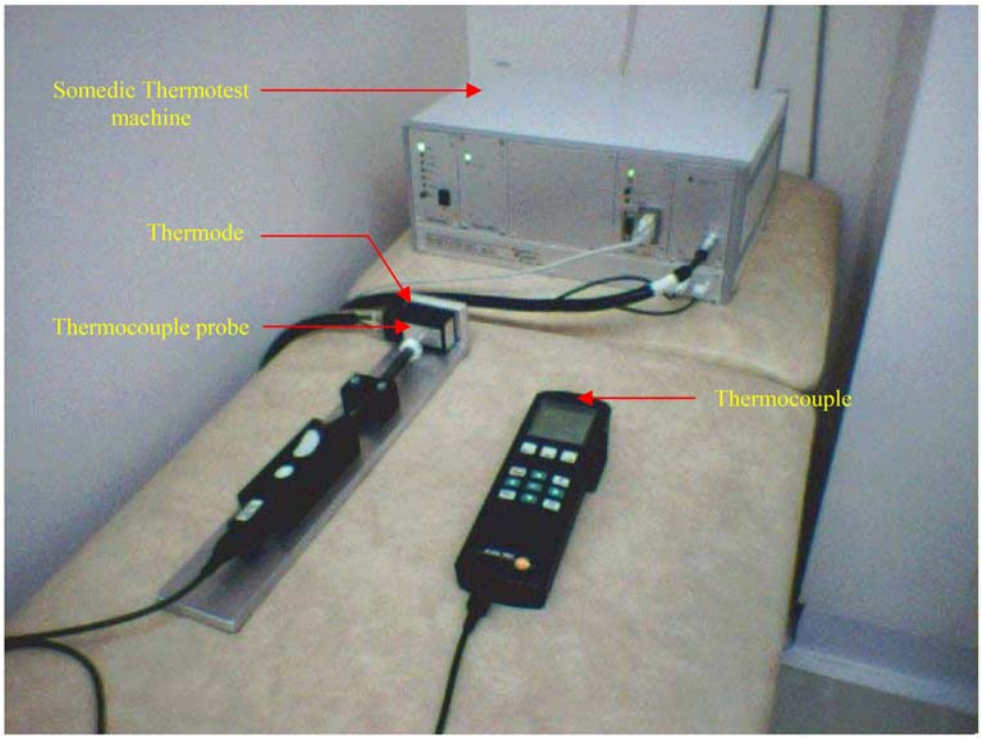

Figure 5.1A. The equipment set up used for both the temperature measurement of the thermode and the calibration of the Thermotest. The individual components of the measurement and calibration equipment are indicated and labeled.

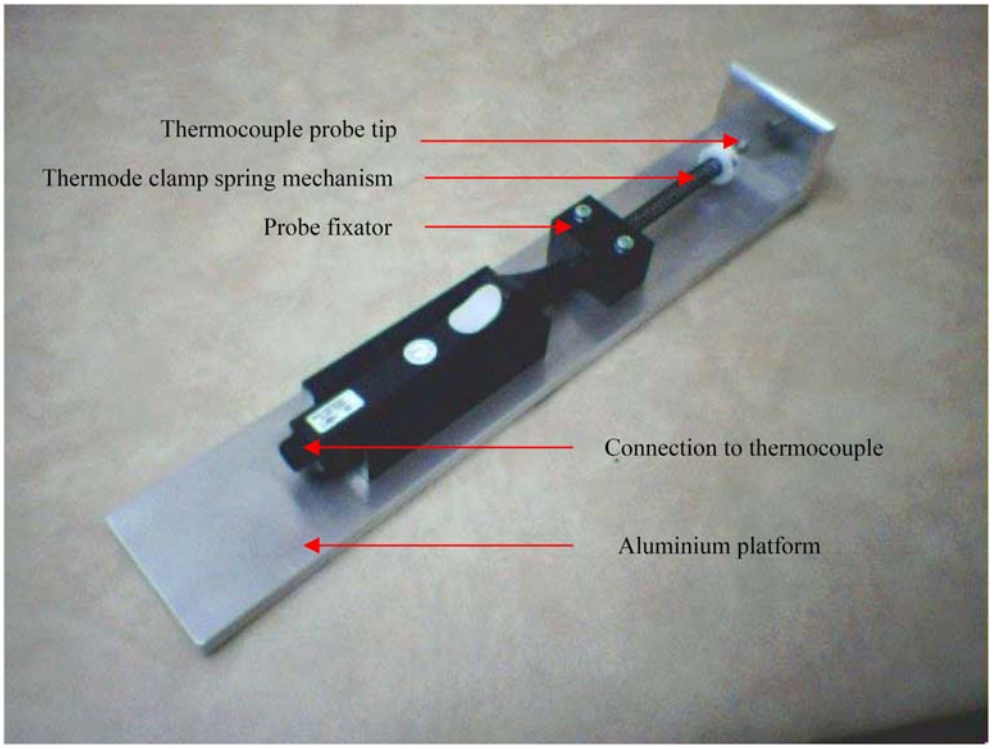

Figure 5.1B. The probe-spring mechanism was used as a clamp for the Thermotest thermode during the study.

probe (Testo A.G., Germany) for temperature measurements was used. The thermocouple has a resolution of 0.001°C and a reported accuracy of 0.05°C. The probe was connected to a spring mechanism within the aluminium frame by the manufacturer of the Thermotest (see Figure 5.1B). The probe-spring mechanism acted as a clamp to hold the thermotest thermode during the experiment. To improve thermal conduction between the surface of the thermode and thermocouple probe, a layer of silicon-based heat transfer compound (Electrolube, HTC10S, United Kingdom) was applied between the contact surfaces.

5.6.3 Pre-calibration thermode temperature measurement procedures

For the purposes of this thesis, the specific temperatures of 44°C, 45 °C, 46 °C, 47°C, 48°C and 49°C were tested. These temperatures were within the temperatures normally considered to stimulate nociceptors associated with noxious thermal stimuli (Lamotte & Campbell, 1978; Meyer & Campbell, 1981; Van Hees & Gybel, 1981; Yarnitsky et al, 1995). The temperatures used also included the temperatures used for inducing noxious stimulation (45°C, 46°C, 47°C, 48°C and 49°C) for all other studies in this thesis. The order of the temperatures tested was randomised using an online pseudo-randomisation algorithm (http://www.randomization.com, accessed 10th Aug 2003). Three measurements were taken for each temperature. But not consecutively. The randomisation of the temperatures was performed in three blocks, each consisting of all six temperatures sampled without replacement. When one block of randomised temperatures testing was completed, the testing for the next block of temperatures was then conducted.

The researcher preset the temperature for the thermode using EXPOSURE. The thermode temperature was allowed to stabilise for 120 seconds (duration recommended by the manufacturer) before the temperature read-out was obtained from the thermocouple. The temperature was recorded before the next randomised temperature was tested. This procedure was repeated for all the temperatures.

5.6.4 Calibration procedures

The calibration procedures were automated using the EXPOSURE software provided by Somedic A.B. Figure 5.2 shows the computer screen-capture of the EXPOSURE software interface. Before calibration, the experimenter specified the range of

temperatures to be calibrated. Figure 5.2 indicates that the temperature range of 44°C to 49°C was programmed to be calibrated (Figure 5.2, Box 1). At the start of calibration, EXPOSURE raised the thermode temperature to 44°C, the lower boundary of the calibrated temperature range. After the required temperature was reached on the thermode, EXPOSURE started the timer for a duration of 120 seconds. This was to allow the thermode temperature to stabilise before a measurement was taken from the thermocouple. At the end of 120 seconds, the temperature was read off the thermocouple display by the researcher. This temperature value was entered into EXPOSURE manually by the researcher. These same procedures were repeated for the upper temperature boundary for the calibrated temperature range. In this experiment, the upper temperature boundary was 49°C. EXPOSURE also required a mid-range value for the calibration. This was automatically chosen by EXPOSURE to be 47°C. The procedure for obtaining the thermode temperatures for the lower and upper boundaries was repeated for the mid-range temperature. Once the new calibration settings have been confirmed, it was saved as a .txt file. This provided a record of previous calibration settings for the trouble-shooting of any equipment problems that might arise in the future.

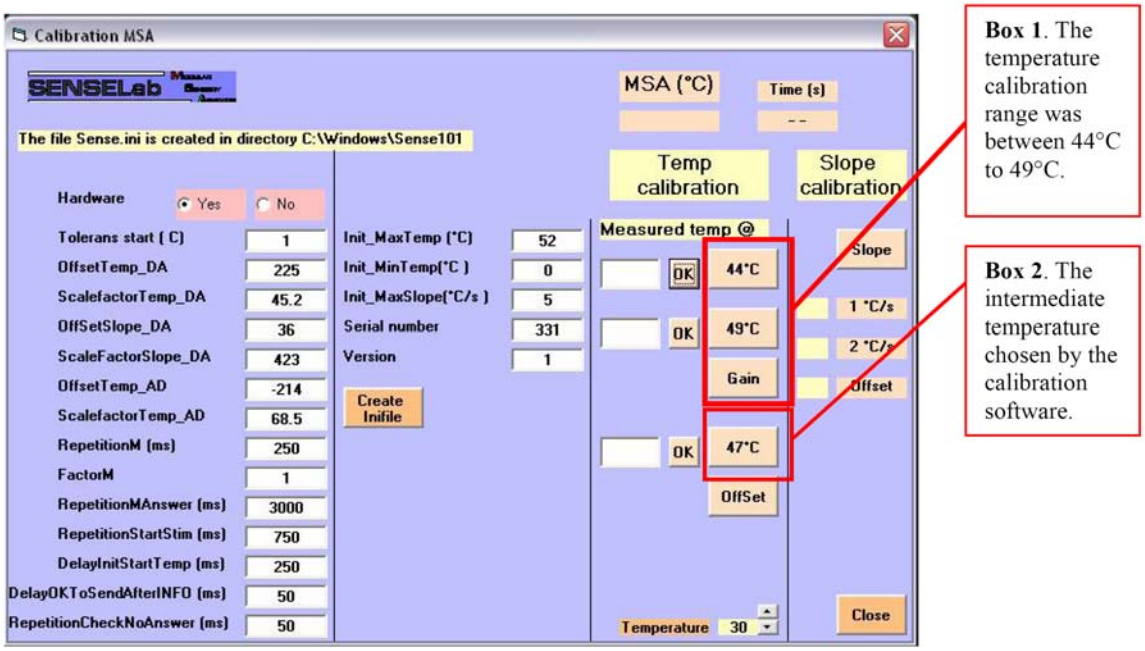

Figure 5.2. Screen capture of the software used to calibrate the Somedic Thermotest machine. Box 1: The lower and upper temperatures determined by the experimenter for calibration. In this example, the lower and upper temperature were 44°C and 49°C respectively. Box 2: The intermediate temperature chosen by the software for the calibration process. In this example, the software selected 47°C.

5.6.5 Post-calibration thermode temperature measurements

Once the calibration was completed, the thermode temperatures in the range of between 44 °C and 49°C were measured again to ascertain the effect of the calibration as in Section 5.6.3.

5.7 Results for Experiment A

5.7.1 Room temperature and length of experiment

The initial room temperature was 22.0°C and the end room temperature was 22.5°C. Experiment A took 2 hours and 39 minutes.

5.7.2 Accuracy and precision of temperature measurements

Table 5.1 shows the mean and standard deviations of the temperatures measured pre- and post-calibration. A graphical representation of Table 5.1 was plotted as Figure 5.3. The Y-axis values on Figure 5.3 are the differences between the mean temperatures and the reference temperatures. The standard deviations are shown as error bars on the graph. The Y-axis was plotted as the differences of mean and reference temperature for ease of interpretation. Intuitively, any coordinates deviating away from y = 0 would allow immediate recognition that the temperatures are either positively deviated (i.e. the temperatures measured are higher than the reference temperatures) or negatively deviated (i.e. the temperature are lower than the reference temperatures).

Table 5.1

Mean Thermode Temperature and Associated Uncertainty of Pre- and Post-Calibration of Somedic Thermotest for Experiment A

	Pre-calibration		Post-calibration	
Reference temperature / °C	mean / °C	standard deviation	mean / °C	standard deviation
44	44.07	0.04	43.99	0.02
45	45.04	0.02	44.98	0.03
46	46.03	0.03	45.97	0.01
47	47.03	0.03	46.96	0.02
48	48.02	0.03	47.95	0.01
49	49.00	0.08	48.96	0.01

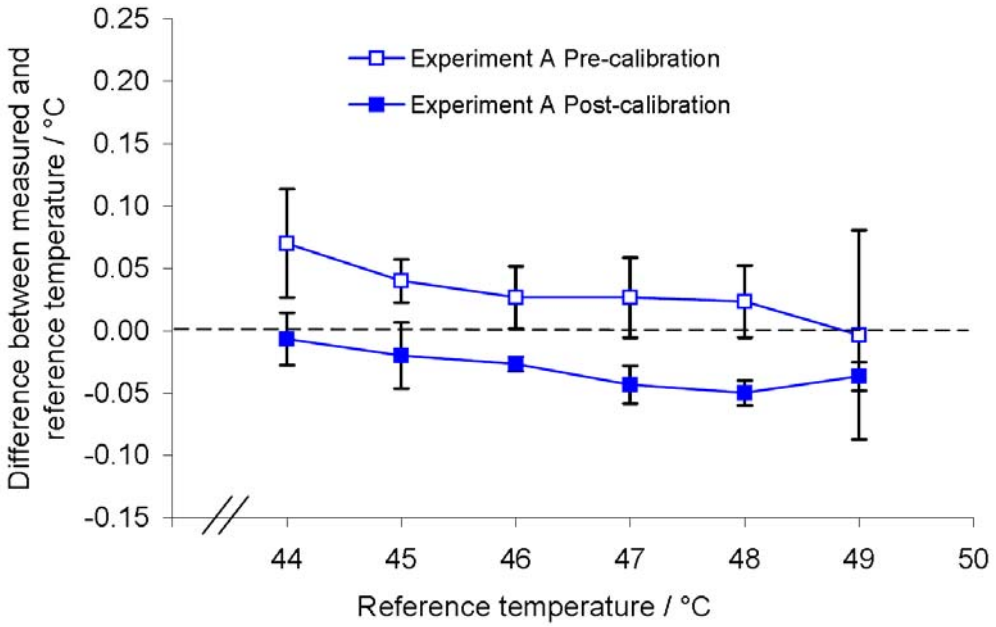

Figure 5.3. The graph showed the difference between the measured and reference temperatures on the thermode surface for the pre-calibration and post-calibration conditions in Experiment A. The temperature measurements of the two conditions differ by the same absolute temperature magnitude compared to the reference temperature. However, the temperature measurements were less variable for the post-calibration condition. The error bars indicate the standard deviation.

Two features were noted from Figure 5.3. Firstly, the pre-calibration temperature measurements overestimated the reference temperatures and the post-calibration measurements underestimated the reference temperatures. However, the mean temperature differences were of approximately the same magnitude between the two calibration conditions. The overall mean temperature difference was +0.03°C for the pre-calibration condition and -0.03°C for the post-calibration condition. Figure 5.3 showed that the graph for the post-calibration condition is roughly a downward transposition of the pre-calibration graph.

Secondly, most of the post-calibration temperatures had smaller standard deviations compared to the pre-calibration temperatures. In terms of repeatability, the post-calibration measurements were less variable than the pre-calibration measurements.

The mean standard deviations of all the temperatures were 0.04°C for the pre-calibration temperatures and 0.02°C for the post-calibration temperatures.

5.8 Discussion for Experiment A

Two issues concerned with this experiment were investigated: the accuracy and precision of the temperature measurements. Practically, the accuracy of measurements is seldom ascertained in an experiment. This is because the 'true' value of the particular variable investigated can never be known with certainty. Nevertheless, for this experiment, we have operationally defined the 'true' value of the variable to be the reference temperatures measured. This circumvented the problem by setting the 'true' values to be the temperatures that the experimenter has pre-programmed the machine or software to produce for testing. Once the qualitative concept of accuracy has been operationally defined, the comparison of the accuracy of measurements can be conducted. The results from Table 5.1 and Figure 5.3 showed that the post-calibration temperature measurements were not more accurate than the pre-calibration measurements. The absolute magnitude of temperature differences between the measured and reference temperatures was 0.03°C. This difference can be considered to be practically negligible.

Regarding the precision of the measurements, the post-calibration temperature measurements were more precise than the pre-calibration measurements. The precision was represented by the standard deviations of the temperature measurements. In this experiment, the post-calibration measurements were less variable than the pre-calibration variables. This was especially notable for the pre- and post-calibration temperatures for the reference temperatures of 46°C and 49°C. The post-calibration temperature variability decreased from a pre-calibration SD value of 0.04°C to 0.02°C for the reference temperature of 44°C, and 0.08°C to 0.01 for the reference temperature of 49°C.

Although pre-calibration temperature measurements had a maximum standard deviation of 0.08°C, it is quite unlikely that this will have any practical impact on experiments involving human subjects judging temperature magnitudes. The temperature variability of 0.08°C is relatively small compared to the capacity of

human observers' discrimination ability that ranges between 0.2°C to 0.3°C (LaMotte & Campbell, 1978; Bushnell, Taylor, Duncan & Dubner, 1983; Robinson, Torebjork, & LaMotte. 1983). In other words, although the post-calibration temperatures were relatively accurate and precise, even in the uncalibrated state, the temperature variability was small enough not to influence the judgements of human observers.

5.9 Summary

The calibration procedure improved the precision of the post-calibration temperature measurements. This meant that the variability of the temperature measurements post-calibration was reduced. The calibration procedure also shifted the temperature measurements below the reference temperatures with a mean absolute difference of 0.03°C. However, the overall magnitude of temperature measurement accuracy was not affected.

Experiment B

5.10 Aims

a. To estimate the accuracy and precision of the temperatures produced by the Thermotest for the pre-calibrated and post-calibrated states.
b. To determine the repeatability of the temperature measurements between the post-calibrated state of Experiment A and pre-calibrated state of Experiment B.

5.11 Methods

The experiment was conducted in the same room as for Experiment A. The same equipment, measurement and calibration procedures were also used. The Somedic thermotest was not calibrated at any point over the 6 months period between Experiments A and B.

5.12 Results

5.12.1 Room temperature and length of experiment

The initial room temperature was 22.0°C and the end room temperature was 23.0°C. The whole experiment took 1 hour and 23 minutes. Experiment B took a shorter time to complete compared to Experiment A due to the experimenter's increased familiarity with the calibration equipment.

5.12.2 Temporal drift of temperature measurements

The post-calibration temperature measurements from Experiment A were used to compare with the pre-calibration temperature measurements from this experiment. This was to ascertain any changes in temperature measurements due to time (temporal drift). The means and standard deviations of the temperature measurements of Experiment A post-calibration and Experiment B pre-calibration conditions are shown in Table 5.2. Figure 5.4 is the graphical representation of Table 5.2. The Y-axis represents the differences between the mean temperatures and the reference temperatures. The standard deviations are represented by the error bars on the graphs.

Table 5.2
Mean Thermode Temperature and Associated Uncertainty of Experiment A Post- and Experiment B Pre-Calibration of Somedic Thermotest

	Experiment A Post-calibration		Experiment B Pre-calibration	
Reference temperature / °C	mean / °C	standard deviation	mean / °C	standard deviation
44	43.99	0.02	43.94	0.07
45	44.98	0.03	44.95	0.06
46	45.97	0.01	45.93	0.04
47	46.96	0.02	46.91	0.01
48	47.95	0.01	47.91	0.02
49	48.96	0.01	48.91	0.02

Figure 5.4 demonstrated that there was a temporal drift of the Experiment B pre-calibration temperature measurements towards a negative difference of the measured and reference temperatures. Using Experiment A measurements as the baseline in this

instance, there was an overall mean decrease of 0.06°C from Experiment A post-calibration condition to the Experiment B pre-calibration condition.

Over the period of 6 months, there was an increased variability in the temperature measurements. This was especially evident for the temperatures of 44°C, 45°C and 46°C. The standard deviations increased from 0.02°C to 0.07°C, 0.03°C to 0.06°C, and 0.01°C to 0.04°C for the reference temperatures of 44°C, 45°C and 46°C respectively (see Table 5.2).

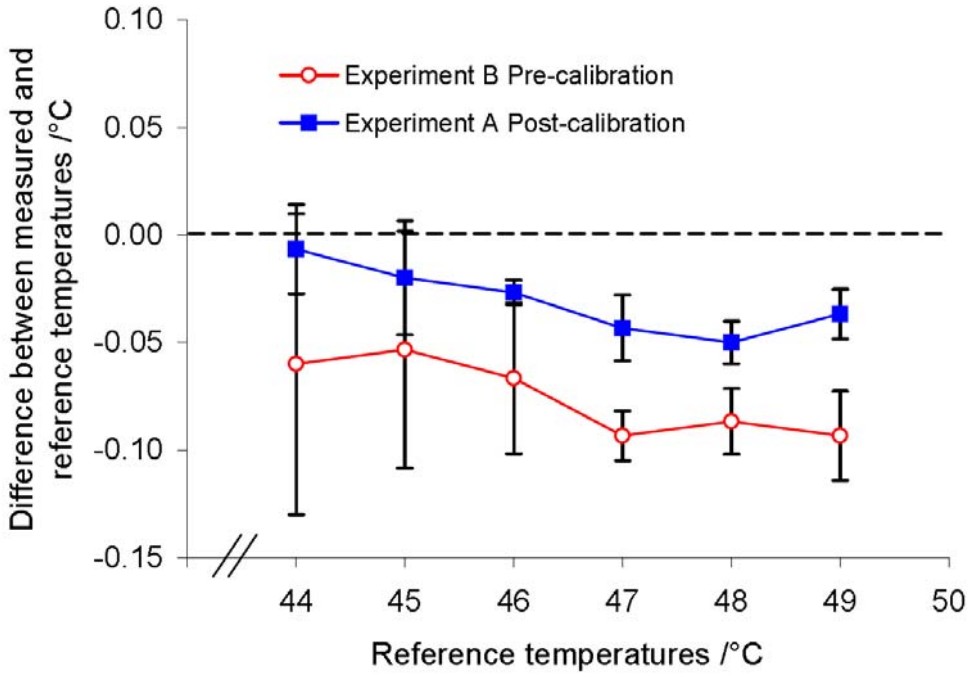

Figure 5.4. The graph showed a downward shift of the measured temperatures from the thermode within a 6 month period. This is evidence for a drift of the temperatures produced by the thermode. There is also an increase in the variability of the temperature measurements in the Experiment B pre-calibration condition. The error bars indicate the standard deviation.

5.12.3 Accuracy and precision of Experiment B temperature measurements

The mean temperature measurements and associated standard deviations are shown in Table 5.3. Figure 5.5 is the graphical representation of Table 5.3. The Y-axis in Figure 5.5 again represented the difference between the mean measured temperatures and the reference temperatures.

As seen in Figure 5.5, the calibration procedure has corrected the temporal temperature drift of the Thermotest by raising the mean measured temperatures by

approximately 0.11°C. The largest correction was for the reference temperature of 49°C, with a change from a mean temperature of 48.91°C to 49.05°C.

The calibration also decreased the variability of the measured temperatures. This was evident from the smaller error bars on the graph representing the post-calibration values. The mean standard deviation for the pre-calibration was 0.04°C. This was decreased to 0.02°C after the calibration procedure.

Table 5.3

Mean Thermode Temperature and Associated Uncertainty of Pre- and Post-Calibration of Somedic Thermotest for Experiment B

	Pre-calibration		Post-calibration	
Reference temperature / °C	mean / °C	standard deviation	mean / °C	standard deviation
44	43.94	0.07	44.03	0.01
45	44.95	0.06	45.04	0.01
46	45.93	0.04	46.04	0.01
47	46.91	0.01	47.03	0.02
48	47.91	0.02	48.02	0.02
49	48.91	0.02	49.05	0.04

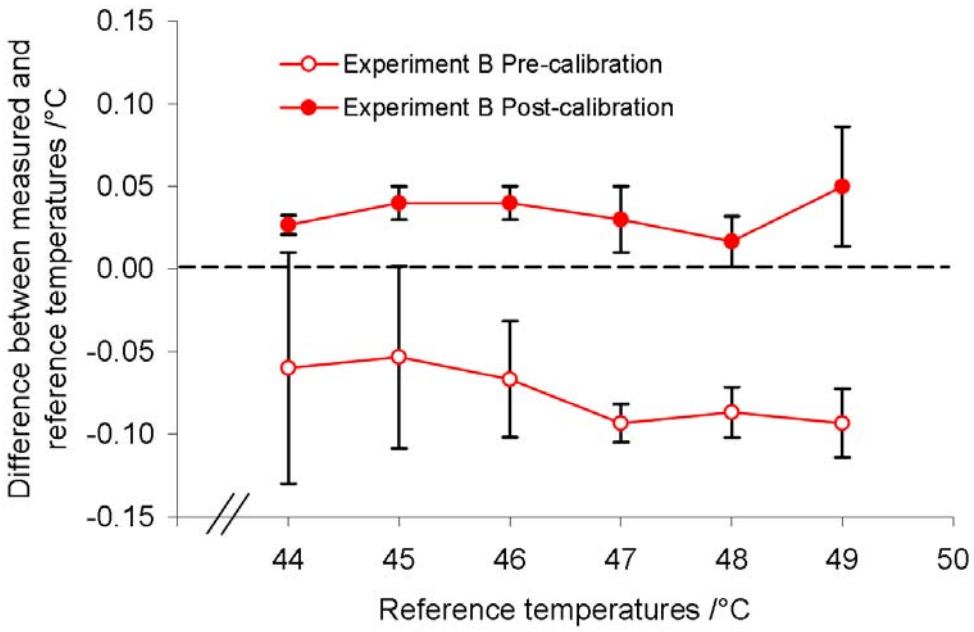

Figure 5.5. The graph showed the difference between the measured and reference temperatures on the thermode surface for the pre-calibration and post-calibration conditions in Experiment B. After the calibration, the average temperature measurements move upwards and closer towards the reference temperature. There was also a decrease in the variability of the post-calibration measurements. The error bars indicate the standard deviation.

5.13 Discussion

The temporal drift of the temperatures demonstrated the importance of periodic checks and examination of the Somedic Thermotest. For the time duration of six months between Experiment A and Experiment B, there was a mean decrease of 0.06°C for the temperature range of 44°C - 49°C. Considering the aforementioned finding that the smallest temperature change discriminable by human observers was 0.20°C, the magnitude of temperature decrease observed in this experiment could be deemed to be relatively small. However, if left unchecked for relatively long periods of time, this could have an impact on the temperature measurements through larger drifts.

Accuracy was not the only concern if the equipment was not monitored on a regular basis. Precision, the variability of the measured values, was an important factor to be considered for the efficient functioning of the equipment. As was noted in Figure 5.4, the variability of the temperatures produced by the Thermotest increased over the period of six months. With greater variability of temperature production, this would cause more uncertainty in the temperatures produced by the equipment. The temporal drift and increased variability was corrected by the calibration procedure (see Figure 5.5). The decreased variability of temperature measurements from 0.04°C to 0.02°C was consistent with the post-calibration variability of measurements from Experiment A. This indicated that monitoring of the equipment's accuracy and precision in producing the thermode temperatures should be performed periodically. This is to ensure that the accuracy and precision of the temperature measurements are within the specifications required by the study.

5.14 Summary

For the inter-experimental time period of 6 months, there was a temporal drift in the temperature measurements from Experiment 1 to 2. There was also an increase in variability of the thermode temperatures. However, this was corrected by the calibration procedure. The calibration reduced the discrepancy between the mean measured temperatures and reference temperatures. It also decreased the variability of the temperatures' measurements. It was recommended that checks should be

performed on the Somedic Thermotest before the conduct of a new study to prevent inaccuracies and imprecision in the temperature production.

Experiment C

5.15 Aims

a. To estimate the accuracy and precision of the temperatures produced by the Thermotest for the pre-calibrated and post-calibrated states.
b. To determine the reproducibility of the temperature measurements between the post-calibrated state of Experiment B and pre-calibrated state of Experiment C.
c. To determine the reproducibility of the temperature measurements between the test site of Experiments A and B to Experiment C.

5.16 Methods

5.16.1 New test environment

The experiment was conducted at a new test site. This test site was located in a room at a Western General Hospital, Edinburgh. The volume of the L-shaped room was approximately 9.61m^3. Therefore the volume of the room was much smaller than the previous test site by a difference of 26.29m^3.

The test site was neither sound-proof nor temperature controlled. However, minimal sound disturbance from sources external to the room was heard when the door to the room was closed. The room was poorly ventilated. This posed potential problems of greater ambient temperature variability when the room was used for several hours or during the summer months. Since the room temperature was not controllable by any simple means, the ambient temperature was monitored via a mercury thermometer throughout this experiment.

5.16.2 Temperature measurements and calibration procedures

The temperature measurements were obtained using similar procedures in Experiment A (see Sections 5.6.2 – 5.6.4). The calibration procedure was also similar (see section 5.6.4).

5.16.3 Inter-experiment time period

The inter-experiment time period between Experiments B and C was 9 months. This was planned to coincide with the conduct of the study based at the new test site, the Western General Hospital Edinburgh (Chapter 9).

5.17 Results for Experiment C

5.17.1 Room temperature and length of experiment

The initial room temperature was 24.5°C and the end room temperature was 25.0°C. Experiment C took 1 hour and 25 minutes. This duration was similar to Experiment B.

5.17.2 Changes in temperature measurements between Experiment B post- and Experiment C pre-calibration values

The mean temperature measurements and the standard deviations are shown in Table 5.4. Figure 5.6 is the graphical representation of Table 5.4. The Y-axis values on Figure 5.6 are the differences between the mean temperatures and the reference temperatures. The standard deviations are shown as error bars on the graph.

Table 5.4

Mean Thermode Temperature and Associated Uncertainty of Experiment B Post- and Experiment C Pre-Calibration of Somedic Thermotest

	Experiment B Post-calibration		Experiment C Pre-calibration	
Reference temperature / °C	mean / °C	standard deviation	mean / °C	standard deviation
44	44.03	0.01	43.89	0.03
45	45.04	0.01	44.88	0.03
46	46.04	0.01	45.88	0.06
47	47.03	0.02	46.86	0.10
48	48.02	0.02	47.85	0.04
49	49.05	0.04	48.84	0.04

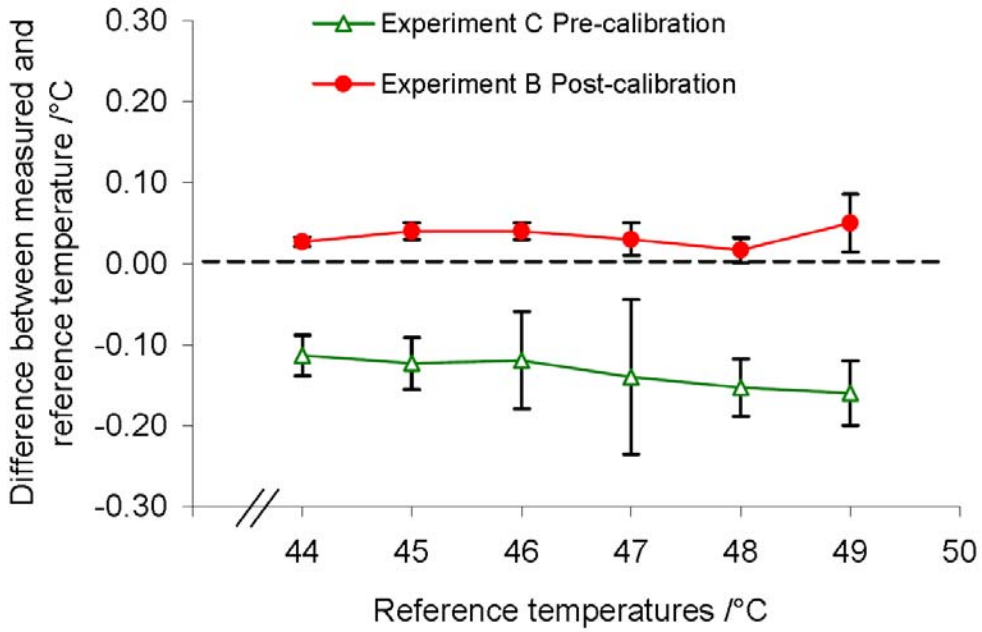

Figure 5.6. The graph showed an downward shift of the measured temperatures for an inter-experimental period of 9 months. This shift was in the same direction of that recorded between Experiment A and B (see Figure 5.4). There was also an increased variability in the temperatures for the Experiment C pre-calibration condition. The error bars indicate the standard deviation.

As seen in Figure 5.6, the mean temperature measurements for the Experiment C pre-calibration temperature measurements have deviated more negatively away from the reference temperatures. There was an overall mean change of -0.16°C from the Experiment B post-calibration temperature measurements to the Experiment C pre-calibration measurements.

There was also an increase in the variability of the temperature measurements. The mean increase in standard deviation for the temperature range of 44°C to 49°C was 0.03°C. The two largest increases in standard deviation were for 46°C and 47°C with a change of 0.05°C and 0.08°C respectively.

5.17.3 Comparison of measurements between Experiment C pre- and post-calibration values

Table 5.5 shows the mean temperature measurements and their standard deviations. Figure 5.7 is the graphical representation of Table 5.5.

Table 5.5

Mean Thermode Temperature and Associated Uncertainty of Experiment C Pre- and Post-Calibration of Somedic Thermotest

	Pre-calibration		Post-calibration	
Reference temperature / °C	mean	standard deviation	mean	standard deviation
44	43.89	0.03	43.99	0.02
45	44.88	0.03	45.01	0.01
46	45.88	0.06	46.00	0.01
47	46.86	0.10	46.97	0.02
48	47.85	0.04	47.97	0.02
49	48.84	0.04	49.98	0.02

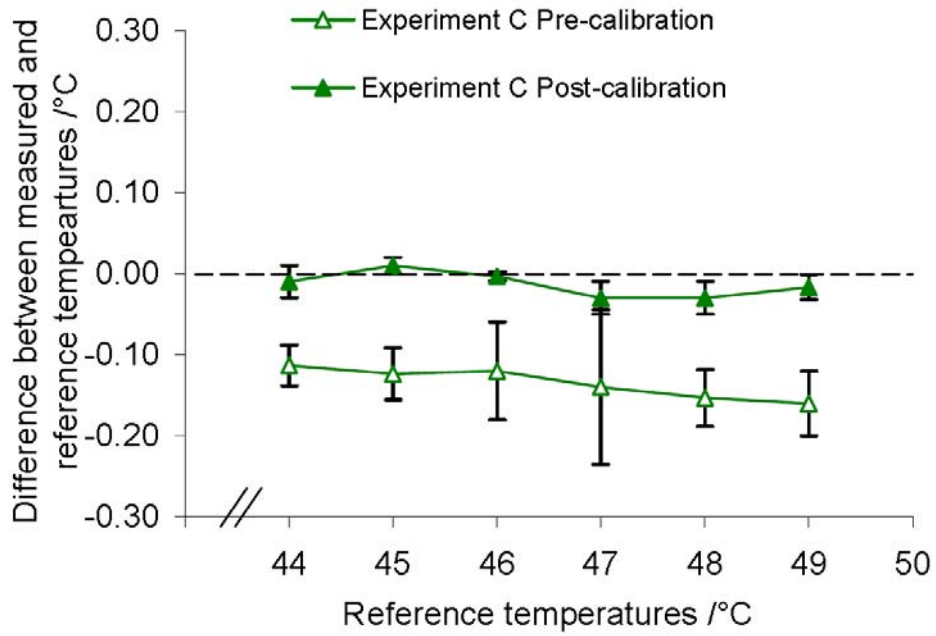

Figure 5.7. The graph showed the difference between the measured and reference temperatures on the thermode surface for the pre-calibration and post-calibration conditions in Experiment C. After the calibration, the average temperature measurements move upwards and closer towards the reference temperature. There was a considerable decrease in the variability of the post-calibration measurements. The error bars indicate the standard deviation.

The calibration procedure shifted the temperature measurements closer towards the reference temperatures. The average change in temperature for the tested range was 0.13°C. The two largest increases in temperature were for 45°C and 49°C with temperature increases of 0.13°C and 0.14°C respectively.

The variability of the pre-calibration measurements was larger than the post-calibration measurements. The mean standard deviation of the pre-calibration measurements was 0.05°C compared to 0.02°C for the post-calibration measurements. The two most variable pre-calibration temperature measurements were for the 46°C and 47°C with standard deviations of 0.06°C and 0.10°C respectively.

5.18 Discussion

5.18.1 Temperature changes between Experiment B to Experiment C

In Experiment C, there were several changes to the testing condition compared to Experiment B. The changes were the test location, the room volume and the ambient temperature. All these factors could have potentially contributed to the observed changes in temperature measurements. As discussed previously, the Thermotest temperature changes were practically trivial because it would not produce a noticeable change in sensation felt by a human observer. Also, calibration procedures would have easily corrected the temperature changes and narrowed the variability of the temperatures.

5.18.2 The effect of the calibration procedures

Similar to previous experiments within this study, the measured temperatures for this study were closer to the reference temperatures after the calibration procedure. An examination of Figures 5.4, 5.5 and 5.6 showed that the calibration procedures brought the measured temperatures within $\pm$ 0.05°C of the reference temperatures. This meant that temperatures produced by the Somedic Thermotest were sufficiently accurate for sensory testing purposes.

Any increased variability of the temperatures was also controlled through the calibration procedures. From the results of Experiments A to C, the variability usually

decreased from the pre-calibration mean standard deviation value of 0.04°C to the post-calibration mean standard deviation of 0.02°C. This predictable reduction in variability could be used as a form of standard for future calibration attempts. Future calibrations which yield mean standard deviations of more than 0.02°C could be deemed unsatisfactory and another calibration should be attempted.

5.19 Summary

The temperature measurement changes in Experiment C deviated positively away from the reference temperatures. Two of the factors that might have contributed to the temperature measurement changes were physical characteristics and the change of the testing location.

The calibration in Experiment C corrected the temperature measurement changes towards the reference temperatures and decreased the variability of the temperature measurements. The decrease in variability was similar to Experiment A and B. The standard deviation of the temperature produced by the calibrated equipment was approximately 0.02°C.

Experiment D

5.20 Aims

- To determine the relative efficiency of the procedure using a high trial number compared to the procedure using a low trial number
- To examine if the variabilities of the discrimination ability measure for the two trial numbers are statistically significantly different.

5.21 Methods

5.21.1 Participants

Participants were recruited from the students and staff of Queen Margaret University, Edinburgh (QMU) using convenience sampling. Requests to potential volunteers for study participation were made through two methods. The first method was recruitment of students and staff by word of mouth within the Physiotherapy Subject

Area, QMU. The second method involved the researcher advertising the study to physiotherapy undergraduate and postgraduate students at the beginning of their classes. Six healthy volunteers (4 men and 2 women) took part in the experiment. The median age of the participants was 21.5 years (range: 19-36 years). All participants were right-handed. No participants had prior experience of the experimental protocol.

5.21.2 Ethical approval

This study was approved by Queen Margaret University's Research Ethics Committee. All the participants provided written informed consent for participation in this experiment.

5.21.3 Inclusion and exclusion criteria

The inclusion criteria were: a) age of 18 years or more, b) ability to provide consent for participation in the study. The exclusion criteria were: a) the presence of medical conditions that caused anaesthesia to the tested limb, or the consumption or application of medication that caused analgesia or anaesthesia on the tested limb, b) any wounds or injury to the tested limb; as reported by the participants or observed directly by the experimenter. No participant was excluded on these criteria.

5.21.4 Choice of trial numbers

Two stimulus presentation numbers were chosen based on a literature review of signal detection theory (SDT) studies that investigated thermal pain perception between the periods of 1969-2002 (see Table 5.6). These numbers were the median stimulus numbers of studies that used parametric and non-parametric SDT measures. The stimulus numbers were 40 trials (range= 6-100) and 17 trials (range= 8-67) respectively. The assumption was studies that utilised parametric measures used relatively higher stimulus numbers in order to meet the statistical requirements for the use of parametric statistics. The stimulus numbers chosen were considered representative of the upper-bound and lower-bound median frequencies used for past studies.

Table 5.6
SDT studies in pain perception using parametric and distribution-free sensitivity indices

Studies using parametric sensitivity index

Author/s	Modality	Trials per intensity	Sensitivity index
Clark (1969)	Radiant heat	25	*d′*
Clark and Mehl (1971)	Radiant heat	16	*d′*
Clark and Mehl (1973)	Radiant heat	15	*d′*
Clark and Dillon (1973)	Radiant heat	25	*d′*
Clark (1974)	Radiant heat	6	*d′*
Clark and Goodman (1974)	Radiant heat	12	*d′*
Clark and Yang (1974)	Radiant heat	24	
Chapman et al (1975)	Electrical to teeth	75	*d′*
Harkins and Chapman (1976)	Electrical to teeth	100	d_e
Harkins and Chapman (1977)	Electrical to teeth	100	d_e
Chapman and Butler (1978)	Electrical to teeth	50	d_e
Rollman (1979)	Electrical to skin	63	d_e
Goolkasian (1980)	Radiant heat	40	d_e
Goolkasian (1983)	Radiant heat	60	d_e
Rollman (1983)	Electrical to skin	40	*d′*, d_e, *Δm*, *P(A)*, *D(A)*
Schumacher and Velden (1984)	Electrical to skin	40	d_s
Dworkin et al (1995)	Radiant heat	16	*d′*
Murphy et al (2002)	Electrical to intramuscular	50	*d′*

Studies using distribution-free sensitivity index

Author/s	Modality	Trials per intensity	Sensitivity index
Chapman, Chen and Bonica (1977)	Electrical to teeth	67	*A*
Callaghan et al (1978)	Electrical to skin and radiant heat[1]	12	*E*
Yang et al (1979)	Radiant heat	14	*P(A)*
Naliboff and Cohen (1981)	Radiant heat	24	*A′*
Clark et al (1981)	Radiant heat	20	*P(A)*
Cohen (1983)	Radiant heat	4	*P(A)*
Janal et al (1984)	Radiant heat	12	*P(A)*
Yang et al (1985)	Radiant heat	8	*P(A)*
Lautenbacher et al (1989)	Contact heat	Exp 1: 50 Exp 2: 20	*P(A)*
Yang et al (1991)	Radiant heat	8	*P(A)*
Fuller and Robinson (1993)	Radiant heat	24	*P(A)*
Janal et al (1994)	Radiant heat	12	*P(A)*
Glusman et al (1996)	Contact heat	12	*P(A)*
Kemperman et al (1997)	Radiant heat	40	*P(A)*

5.21.5 Apparatus and stimuli

The set up of the equipment for all studies in this thesis is described in Section 6.7.4 and shown in Fig. 6.1A of Chapter 6. A height adjustable plinth was placed at an appropriate height so that the participant's forearm can be rested comfortably on the thermode. The laptop running the EXPOSURE software was always facing away from the participant so that the screen was not visible. The Thermotest (Somedic AB, Sweden) was used to administer heat stimuli via a contact thermode (with surface measuring 25mm × 50mm). The thermal stimuli were applied on the ventral surface of both forearms. The stimulus sets (45°C, 46°C, 47°C and 48°C) were pre-programmed using the EXPOSURE software (Somedic AB, Sweden).

5.21.6 Procedures

The study consisted of two sessions: one session presented the stimuli at 17 trials per stimulus intensity (N_{17}) and one session presented stimuli at 40 trials per stimulus intensity (N_{40}). Within each session, 3 noxious temperature-pairs (45°C & 46°C, 46°C & 47°C, 47°C & 48°C) were tested. Each temperature-pair was consigned to one block of testing. The 2 sessions, 3 blocks and trial sequences within the blocks were randomised. All randomisations within this experiment were performed using an online randomisation plan generator (http://www.randomization.com, accessed 4th Feb 2004).Within each block, the experimenter tested subjects' ability to discriminate between two thermal intensities. Contact thermal stimuli were administered via the Thermotest (Somedic AG). The thermal stimuli were applied onto the subjects' dominant forearm using a Peltier contact thermode. Each trial lasted 3 seconds. After each trial, subjects rated their confidence of whether the higher or lower intensity of the temperature-pair was presented. The rating response set (confidence rating scale) used by the subject is shown in Figure 5.8. After each response, the subject was provided feedback on the actual temperature presented.

1	2	3	4	5	6
I am absolutely certain the weak stimulus was presented	I am fairly certain the weak stimulus was presented	I am somewhat certain that the weak stimulus was presented	I am somewhat certain that the strong stimulus was presented	I am fairly certain the strong stimulus was presented	I am absolutely certain the strong stimulus was presented

Figure 5.8. The confidence-rating scale was presented to the participant for judgment. representing the discrimination method. The participant verbally provided the number that matched the description of their degree of confidence about which of two stimuli (stronger or weaker) was presented.

5.22 Results

5.22.1 Descriptive statistics

The signal detection theory index, *d′*, was used to represent discrimination ability. Figure 5.9 shows the subjects' discrimination ability for the 3 temperature-pairs. The N_{40} *d′* tended to decrease as the temperatures increased. However this trend was not evident for the N_{17} *d′*. No statistically significant linear trend were found for the within-subject effect of temperature-pairs ($F(1,5) = 0.159$, $p = 0.706$) and the interaction between trial numbers and temperature-pairs($F(1,5) = 1.576$, $p = 0.265$).

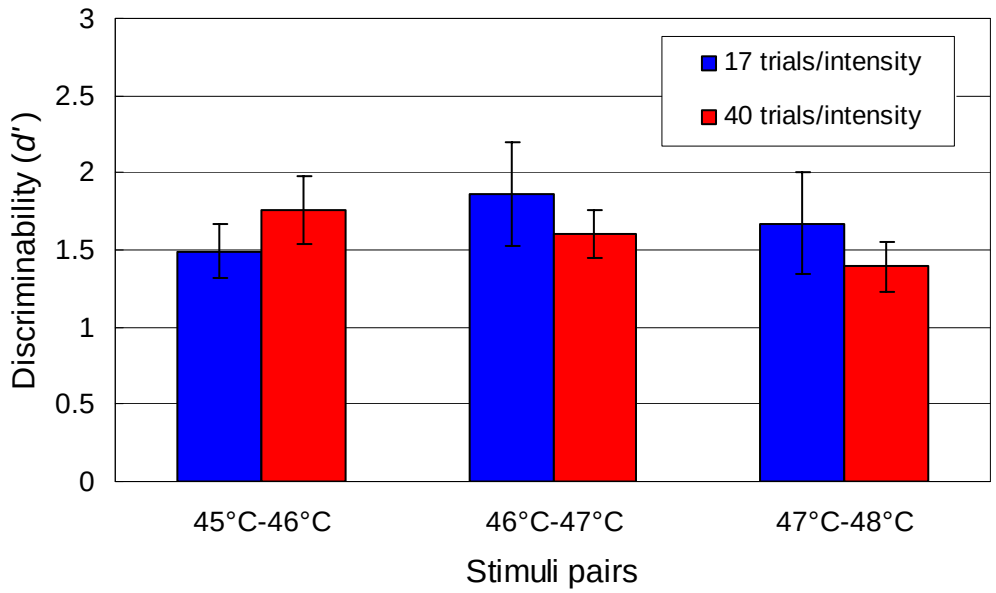

Figure 5.9. The discriminability of the stimuli pairs and trial number conditions. There is no statistically significant difference of the discriminability obtained for both trial number conditions. The error bars represent standard error.

5.22.2 Relative efficiency and comparison of variances

The statistical efficiency of a procedure is defined as the dispersion of the sampled values of the measure around the expected mean. A procedure is relatively more efficient than another procedure when it has a smaller dispersion of the sampled values around the expected mean. The *d′* variances of the N_{17} and N_{40} were var(N_{17}) = 0.513 and var(N_{40}) = 0.295, respectively. Therefore the relative efficiency N_{40} compared to N_{17} is 1.74.The N_{40} procedure was relatively more efficient than the N_{17} procedure from the variance values.

The Levene's test for equality of variances was performed to compare if there is statistical significant difference between the variances for N_{17} and N_{40}. It was found that there was no statistically significant difference between the variances of N_{17} and N_{40} ($F(1,34) = 1.470$, $p = 0.234$).

5.22.3 Inferential statistics

A 2 x 3 (stimulus numbers × temperature-pairs) repeated measures ANOVA test was used to analyse the data. There were no significant main effects between conditions N_{40} and N_{17} ($F(1,5)= 0.454$, $p = 0.531$). No significant interaction was found between the trial numbers and temperature-pairs used ($F(2,10) = 0.934$, $p = 0.425$). Post-hoc pairwise comparisons between all the temperature-pairs' *d′* for all conditions showed no statistically significant comparisons.

5.23 Discussion

It was found that N_{40} procedure was relatively more efficient than N_{17} procedure. This finding was consistent with Hautus' (1997) and Miller's (1996) results. Their study found that when the stimulus number was decreased, the variability of the discriminative measure increased, hence relative efficiency decreased. The implication of this finding was that a trade-off should be found between the stimulus number and the variability of the procedure. However, there were no statistically significant differences between the variances of discriminability obtained from the two trial numbers. This study provided some evidence that a lower stimulus number may provide an acceptable estimate of the discrimination measure.

The number of subjects employed in an experiment was a factor not considered in our study but examined in Hautus' (1997) and Millers' (1996) study. Under the circumstance of unacceptable variability within the discriminative measure due to low stimulus number, the use of higher number of subjects and data pooling may reduce the variability of the measure (Hautus, 1997).

5.24 Summary

No statistically significant difference was found for the discrimination measure and the variances between N_{17} and N_{40}. The use of a lower stimulus number (i.e. 17 trials per stimulus intensity) in our study will not cause extreme deviation in the discrimination measure.

5.25 Overall conclusion

The Somedic Thermotest produced thermal stimuli that were accurate and precise. The physical parameters of the thermal stimuli were more than sufficient for the experimental protocol conducted within this thesis. However, the accuracy and precision of the temperatures produced decreased with time and possibly with a change in testing location. It is therefore recommended that diagnostics testing be performed on the equipment before the start of any new experimental work. Also, the use of a lower trial number (i.e. 17 trials per intensity) may be considered acceptable. This conclusion was made based on the statistically non-significance of the comparison between the means and the variances of two experimental condition with different trial number (N_{17} and N_{40}).

Chapter 6
A Common Analytical Framework for Two Rating Methods

6.1 Introduction

Most signal detection theory (SDT) studies in pain perception have used the magnitude-rating scale, as outlined in Chapters 3 and 4. This thesis suggested, in Chapter 4, Section 4.4.2, that the interpretational ambiguity of SDT measures when used in SDT pain perception studies is partly due to the interpretation of results obtained using the magnitude-rating scale. This thesis also suggested that this problem of interpretational ambiguity may be resolved through the use of the confidence-rating scale instead of the magnitude-rating scale. There is a possibility the results generated from the magnitude-rating scale and confidence-rating scale may not be comparable (Rollman, 1983). To investigate whether these two scales are comparable, Irwin & Whitehead (1991) and Irwin et al. (1994) used an extension of SDT as a common analytical framework for the data obtained from these scales. The index utilised for estimating the discriminability from both methods was d'. The theories underlying the analytical framework were originally proposed by Braida & Durlach (1972) and Laming (1984) for comparing different psychophysical methods in sensory perception. This analytical framework has been adopted in this thesis. The framework acts as a theoretical bridge between the two scales and their associated methods. The following section will outline the theories for the common analytical framework.

6.2 Direct scaling and discrimination methods

Two types of tasks have been mentioned in previous sections: the magnitude-rating task and the confidence-rating task. These tasks can be classified under more general groups of methods, namely magnitude-rating scale under direct scaling methods and confidence-rating scale under discrimination methods. Direct scaling methods measure the participant's perceived or psychological magnitudes of the stimuli directly from the judgments made by the participants (Gescheider, 1997, p396). Some other methods that also fall under this broad category of direct scaling are magnitude estimation, absolute identification and category scaling. Macmillan & Creelman (1991) differentiated between magnitude estimation, absolute identification and category scaling by using a simple system. For each method, the number of stimuli

within the stimuli set and the number of response categories within the response set were compared. If the number of response categories is more that the number of stimuli, then the method is defined as magnitude estimation. If the number of response categories is the same as the number of stimuli, then the method is defined as absolute identification. Lastly, if the number of response categories is less than the number of stimuli, the method is defined as category scaling. Therefore, by definition the magnitude-rating task could be considered a subtype of the magnitude estimation method. This is because the magnitude-rating task, as described in this thesis, consists of more response categories than number of stimuli. In contrast to direct scaling methods, discrimination methods measure the participant's perceived psychological magnitudes of the stimuli from data on the participant's ability to tell one stimulus from another (Gescheider, 1997). One example is the one interval, confidence-rating task described in Chapter 4, Section 4.3.2.

6.3 Influence of stimuli range and response autocorrelation

Laming (1984) argued that the assigned responses to physical stimuli by participants for direct scaling methods may only be considered ordinal. Laming (1984) supported this statement by explaining that previous experiments had shown that participants were only able to reliably assign stimuli to about five categories. This meant the data obtained from direct scaling methods may be analysed by tallying the frequencies assigned to the response categories for each stimulus. The analysis of response frequencies is common to categorical variables. Based on the information obtained from analysis of the ordinal data, it is possible to infer the discriminability of one stimulus from another stimulus for the direct scaling methods. This analysis would produce estimates of discriminability based on magnitude-ratings. However this estimate of discriminability is only approximate. The approximate nature of the discriminability estimate is due to the way stimuli are usually presented in direct scaling methods. In direct scaling methods, several stimuli of different intensities are randomly presented for the participant's judgment. In contrast, the discrimination methods randomly present only two stimuli of different intensities to the participants within a test session. Participants may perform poorer in the direct scaling method because a larger stimuli range is used as compared to the discrimination method. The larger stimuli range in the direct scaling method may lead the participants to compare

the particular stimulus to the lowest and highest stimulus intensity (Parducci, 1965). Or the stimulus could be compared to a particular reference stimulus before a judgment is made (Helson, 1964). These comparisons would in some way alter the judgment provided by the participant. Therefore, the influence of stimulus range is a sminece of variance introduced to the participant's ability to differentiate between stimuli.

Another source of variance that may influence the participant's judgment is the autocorrelation of responses made to previous responses in direct scaling methods. That is, the current response is made based on certain properties of previous responses. The autocorrelation is strongest when the intensities of the current stimulus and the preceding stimulus are very closely spaced (Baird, Green & Luce, 1980; DeCarlo & Cross, 1990; Green, Luce & Duncan, 1977; Jesteadt, Luce & Green, 1977; Laming, 1984).

The influences of stimulus range and autocorrelation of responses reflect the importance of experimental context on the judgments made by participants. These sources of variance are hypothesized to contribute to an overall lower d' for the direct scaling methods compared to the discrimination method (Braida & Durlach, 1972; Irwin & Whithead, 1991; Irwin et al., 1994).

6.4 Framework theory

Braida & Durlach (1972) theorised that the judgment process involved in the discrimination method is generally less cognitively demanding than direct scaling methods. According to the standard model of signal detection theory, the discriminability is:

$$d'_D = \left(\mu_2 - \mu_1\right)/\sigma_D, \tag{6.1}$$

where d'_D is the discriminability between the two adjacent classes of stimuli, μ_1 and μ_2 are the means of the normal probability densities and σ_D is their common standard deviation. When the standard model is extended to encompass the additional

variance (due to either autocorrelation or stimuli range) inherent in the direct scaling method, then:

$$d'_S = (\mu_2 - \mu_1)/(\sigma_D^2 + \sigma_S^2)^{1/2}, \tag{6.2}$$

where d'_S is the discriminability between the two adjacent classes of stimuli in the direct scaling method, σ_D^2 is the stimulus variance associated with the discrimination method, and σ_S^2 is the judgmental variance associated with the direct scaling method (Macmillan & Creelman, 2005, p134). An estimate of the additional judgmental variance relative to the stimulus variance can be obtained by:

$$\sigma_S^2 / \sigma_D^2 = (d'_D / d'_S)^2 - 1 \ , \tag{6.3}$$

(Durlach & Braida, 1969; Macmillan & Creelman, 2005). The derivation of Equation 6.3 from Equation 6.2 is shown in Appendix B.

As mentioned in Section 4.2.2, d' may be regarded as a distance measure, thereby allowing the summation of d' obtained from discrimination of adjacent stimuli. If these separate d' are assumed to be located along a one dimensional continuum, and the means of the distributions associated with these d' are in the order of $\mu_1 < \mu_2 < \mu_3$, where μ_i is the mean of the distribution of the i^{th} stimuli along the continuum, then:

$$d'(1,3) = d'(1,2) + d'(2,3). \tag{6.4}$$

This outcome from the addition of d' obtained from stimuli 1 and 2, and d' obtained from stimuli 2 and 3 is called cumulative d' or cumulative discriminability (Durlach & Braida, 1969; Macmillan & Creelman, 2005, p114). This outcome represents the participant's performance on the psychophysical task for the entire stimuli range. The stimulus difference needed to produce a performance of $d' = 1$ may be obtained from the cumulative discriminability. This stimulus difference is termed the Weber fraction. Weber's fraction is a constant obtained from Weber's law which describes the relationship between the change in stimulus intensity that can just be

discriminated by the participant (ΔT) and the starting intensity of the stimulus (T). The ratio between ΔT and T is a constant, Weber's fraction (k).
Expressed formally,

$$\Delta T = k\,T \quad \text{or} \quad \Delta T/T = k\,.$$

The larger the value for Weber's fraction, the harder it is to differentiate between stimuli for the particular experimental context. Based on Irwin & Whitehead's framework, it is predicted that the Weber fraction for the discrimination method is smaller than the direct scaling methods.

6.5 Initial evidence for the framework

Irwin & Whitehead (1991) extended this common analytical framework for examining direct scaling and discrimination methods using noxious electrocutaneous stimuli. It was found that cumulative d' obtained from direct scaling methods were lower compared to a discrimination method. The mean Weber fraction, an indication of the just noticeable difference in perceived stimulus intensity, for the discrimination method was found to be 0.043. The mean relative variance between identification-discrimination and magnitude estimation-discrimination was 2.22 and 5.37 respectively. Therefore, the additional variance inherent in the direct scaling methods is more than the discrimination method. Irwin et al. (1994) also found that the cumulative d' of the direct scaling methods were lower than the discrimination method. The Weber fraction for the discrimination method and direct scaling method were 0.051 and 0.244 respectively. Rollman (1983) also found that d' obtained from direct scaling methods were generally lower than discrimination methods. These empirical findings satisfy the theoretical predictions of the analytical framework proposed by Irwin & Whitehead (1991).

Rollman (1983) and Irwin & Whitehead's study (1991) used electrical stimuli for the induction of experimental pain. In comparison, the studies in this thesis used noxious thermal stimuli as the source of experimental pain. One reason for the conduct of this experiment is to extend the analytical framework proposed by Irwin & Whitehead (1991) to psychophysical measurements of responses from noxious thermal stimuli. The relevance of noxious thermal stimuli in pain research is established at several

levels. It is the one of the most commonly used physical stimuli used for evoking experimental pain (Gracely, 2006). Neurobiologically, thermal stimuli activate a known narrow range of primary afferent fiber nociceptors, namely C-fiber and Type I and Type II A-fiber nociceptors (Meyer & Campbell, 1981; Treede, Meyer, Raja & Campbell, 1995). In contrast, electrical stimuli bypass receptor transduction and depolarise the Aβ afferents directly. For this reason, electrical stimuli have been considered to be 'unnatural' or 'non-physiological' (Gracely, 2006; Graven-Nielsen, Sergerdahl, Svensson & Arendt-Nielsen, 2001). For the cutaneous application of electrical stimuli, there is the additional challenge of the administered current spreading to adjacent tissues. This may stimulate different receptor field areas for participants with different skin impedance (Graven-Nielsen et al., 2001). At the molecular level, noxious thermal stimuli activate a non-selective cation channel, the transient receptor potential vanilloid-1 (TRPV1) receptor, which is a potential therapeutic target for pharmacological management of pain (Caterina, et al., 1997).

The other reason is to verify previous findings by Braida & Durlach (1972), Irwin & Whitehead (1991) and Irwin et al. (1994) that the direct scaling method produced decreased d', a relatively larger judgment variance and larger Weber fraction compared to the discrimination method. The reasons for the decreased d', larger relative judgment variance and Weber fraction are due to the contribution of larger amounts of variance from a larger range of stimulus intensities judged and autocorrelation of responses in the direct scaling methods.

6.6 Predictions of study

Therefore the predictions for this study, based on Irwin & Whitehead's (1991) framework are:

- Irwin & Whitehead's analytical framework predicts that d' obtained from the discrimination method is larger than the direct scaling method.
- The analytical framework predicts that relative judgmental variance (Equation 6.3) for the direct scaling method over discrimination method is larger than 1. That is, the amount of variance of within the judgments made for the direct scaling method is larger than judgments for the discrimination method.

- The analytical framework predicts that the cumulative discriminability function is larger in value for the discrimination method compared to direct scaling methods. This means that the overall discriminability of participants for the discrimination method is better than direct scaling methods for the entire stimuli range investigated.
- The analytical framework predicts that the Weber fraction for direct scaling methods is larger than the discrimination method. This means that it is more difficult to differentiate between stimuli for the direct scaling methods compared to the discrimination method.
- When all four predictions are met, the results provide strong evidence for Irwin & Whitehead's (1991) framework. This means that d' obtained from the discrimination method and direct scaling methods are comparable.

6.7 Methods

6.7.1 Participants

Participants were recruited from the students and staff of Queen Margaret University, Edinburgh (QMU) using convenience sampling. Requests to potential volunteers for study participation were made through two methods. The first method was recruitment of students and staff by word of mouth within the Physiotherapy Subject Area, QMU. The second method involved the researcher advertising the study to physiotherapy undergraduate and postgraduate students at the beginning of their classes. Six healthy volunteers (4 women and 2 men) took part in the experiment. The median age of the participants was 28 years (range: 21-35 years). All participants were right-handed. No participants had prior experience of the experimental protocol.

6.7.2 Ethical approval

This study was approved by Queen Margaret University's Research Ethics Committee. All the participants provided written informed consent for participation in this experiment.

6.7.3 Inclusion and exclusion criteria

The inclusion criteria were: a) age of 18 years or more, b) ability to provide consent for participation in the study. The exclusion criteria were: a) the presence of medical conditions that caused anaesthesia to the tested limb, or the consumption or application of medication that caused analgesia or anaesthesia on the tested limb, b) any wounds or injury to the tested limb; as reported by the participants or observed directly by the experimenter. No participant was excluded on these criteria.

6.7.4 Apparatus and stimuli

The set up of the equipment for all studies in this thesis is shown in Fig. 6.1A. The height adjustable plinth was placed at an appropriate height so that the participant's forearm can be rested comfortably on the thermode. The laptop running the EXPOSURE software was always facing away from the participant so that the screen was not visible. A Thermotest (Somedic AB, Sweden) was used to administer heat stimuli via a contact thermode (with surface measuring 25mm × 50mm). The thermal stimuli were applied on the ventral surface of both forearms. The positioning of the forearm on the thermode is shown in Fig. 6.1B. The stimulus sets (45°C, 46°C, 47°C and 48°C) were pre-programmed using the EXPOSURE software (Somedic AB, Sweden).

6.7.5 Procedure

There were two tasks: a magnitude-rating task (MRT), representing the direct scaling method, and a confidence-rating task (CRT), representing the discrimination method. Each task was performed on different forearms for each participant, chosen at random without replacement. All randomisations within this experiment were performed using an online randomisation plan generator (http://www.randomization.com, accessed 17th Aug 2003). The participant completed both the MRT and CRT tasks within the same day. Twenty practice trials, similar to the actual trials, were presented at the beginning of each task for familiarisation.

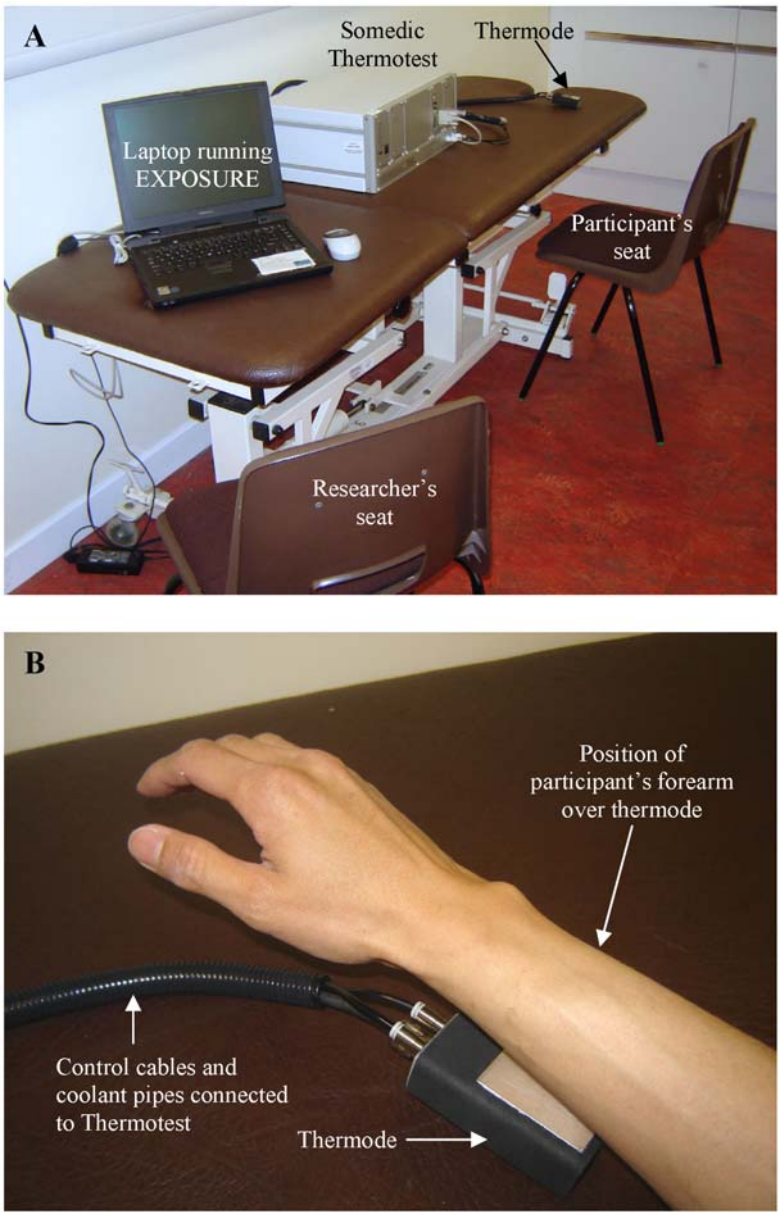

Figure 6.1. A. Equipment set up for studies in this thesis. The laptop screen was not visible to the participant. All equipment rested on a height adjustable plinth so that the thermode is always at a comfortable height for the participant. B. The participant's forearm was comfortably rested on the thermode during the trial. In the figure, the thermode is partly visible to show its position. During an actual trial, the entire thermode is covered by the forearm.

6.7.6 Confidence-rating and Magnitude-rating tasks

The one-interval rating task was used for both CRT and MRT. Each trial began with the experimenter instructing the participant to place his/her forearm on the thermode (pre-set at the relevant testing temperature). A trial contained an observation period of 3 seconds. An automated auditory signal indicated to the participant to remove his/her forearm from the thermode. If the participant was not able to tolerate the full length of stimulus application, he/she was allowed to lift their forearm away from the thermode, although no participants did so during the study. There was an interstimulus interval (ISI) of 10 seconds before the next trial started.

The stimulus set for both tasks consisted of 4 temperatures: 45°C, 46°C, 47°C and 48°C. For the CRT trials, each trial presented one of two temperatures. There was equal probability of presentation for either of the two temperatures. There were three stimulus pairs in total: 45°C and 46°C, 46°C and 47°C, and 47°C and 48°C. The order of stimulus pair presentation was randomised. The three stimulus pairs of the CRT clocked a total of 240 trials per participant (80 trials for each of the three stimulus pairs). For the MRT trials, each trial presented one of four temperatures. Again, there was equal probability of presentation for any one of the four temperatures. There were a total of 160 trials per participant clocked for the MRT (40 trials for each of the four temperatures). Block randomisation, that is randomisation of the four different temperatures within a block, was not performed for the trials. The orders of trial presentation for both tasks were randomised.

The participants verbally indicated their judgments to the experimenter and these were recorded. For both tasks, responses were made based on response sets with six categories. The MRT required the participant to estimate the perceived magnitude of the stimulus presented based on a response set with six descriptions of sensory quality. The six categories are, in increasing magnitude: warm, hot, faint pain, painful, very painful and severe pain (Figure 6.2A). For the CRT, the participant rated their degree of confidence on whether the stimulus presented was the higher or lower intensity of a pair of stimulus intensities (Figure 6.2B).

The participants were told the temperature of the administered stimulus at the end of each trial – that is, trial-by-trial feedback was provided for both tasks. Participants' judgments may be biased by the comparison of observations with a weighted average of stimulus effects. This is also known as the adaptation level effect (Helson, 1964). Feedback was introduced to minimise this bias.

For the MRT, the participants received the following instructions:

> *In this experiment you will be asked to judge the intensities of heat stimuli presented to you. The judgment method involves assigning categories with descriptions to match the intensities of the heat sensations you will experience (Fig. 6.2A shown to the participant). There are six categories of intensities.*

Verbally indicate to the experimenter the category number with a description that matches most closely to the sensation you experienced. After you have done this, you will be told the temperature of the heat stimulus just presented to you.

A

1	2	3	4	5	6
Warm	Hot	Faint Pain	Painful	Very Painful	Severe Pain

B

1	2	3	4	5	6
I am absolutely certain the weak stimulus was presented	I am fairly certain the weak stimulus was presented	I am somewhat certain that the weak stimulus was presented	I am somewhat certain that the strong stimulus was presented	I am fairly certain the strong stimulus was presented	I am absolutely certain the strong stimulus was presented

Figure 6.2. A. Magnitude-rating scale representing direct scaling methods. This scale was presented to the participant during the magnitude-rating task (MRT) for judgment. The participant verbally provided the number that matched the magnitude of sensation felt. B. Confidence-rating scale representing the discrimination method. This scale was presented to the participant during the confidence-rating task (CRT) for judgment. The participant verbally provided the number that matched the description of their degree of confidence about which of two stimuli (stronger or weaker) was presented.

For the CRT, the participants received the following instructions:

In this experiment you will be asked to determine which one of two heat stimuli was presented to you. One stimulus is hotter than the other. Your task is to indicate whether the presented stimulus was the higher or the lower intensity and how confident you are in making that decision. There are six categories to describe your decision (Fig. 6.2B shown to the participant). Verbally indicate to the experimenter the category number with a description

that matches most closely to your decision. After you have done this, you will be told the temperature of the heat stimulus just presented to you.

6.7.7 Prevention of hyperalgesia, heat injury and wind-up

Two specific procedures were implemented to prevent hyperalgesia onset and heat injury of the test sites. The first procedure involved the participant being instructed to shift the position of the thermode to an adjacent forearm skin area at the beginning of a new trial. The second procedure involved the enforcement of an interstimulus interval (ISI) of 10 seconds. This latter procedure also minimised the effect of preceding stimuli increasing the perceived noxiousness of latter trials as a result of temporal summation. This phenomenon of noxious temporal summation is termed wind-up (Price, Hu, Dubner & Gracely, 1977; Staud, Price, Robinson, Mauderli & Vierck, 2004). Wind-up is usually maintained when the ISI is less than 3 seconds. The participant's forearm was checked by the experimenter for signs of heat injury after every 20 trials or if there was a concern that heat injury might have occurred. Signs of heat injury or hyperalgesia, shown by profound erythema with pain or hypersensitivity of the skin, were identified as criteria for withdrawal from the study. No participants suffered any form of heat injury during this study. The entire study duration for each participant was about 2.5 hours.

6.7.8 Analysis

The receiver operating characteristic (ROC) curves of each stimuli pair and task were plotted for every participant. The Gaussian unequal-variance model was fitted to the data using the RScorePlus software written by Lewis Harvey (http://psych.colorado.edu/~lharvey/html/software.html, accessed 24th Oct 2003). RScorePlus is derived from Dorfman & Alf's (1969) RScore program and it provides a maximum-likelihood fit of the signal detection model to the rating data. There were altogether 36 ROCs (6 participants × 3 stimuli pairs × 2 tasks) generated for analysis. For the MRT, the adjacent temperatures were paired for analysis. This yielded the same number of stimulus pairs to the CRT. Data from both tasks were analysed in a similar manner. The detection theory index of discriminability, d_a (Simpson & Fitter, 1973; Macmillan & Creelman, 2005) and the slopes of the ROCs based on three stimulus pairs, s were computed. The index d_a assumes an unequal-variance model and is numerically equal to d' in the equal variance case. The Gaussian equal-variance

index, d' was to be adopted if s for the discrimination and direct scaling data did not systematically depart from unity. One sample t-tests were used to examine if s for the ROCs are significantly different from $s = 1$. The slopes for the linear ROC curves were obtained using the ROC fitting software by Michael Hautus (available at http://www.psych.auckland.ac.nz/people/Hautus/Hautus.htm; accessed 20th Jan 2006). When extreme response frequencies were present (i.e. categories containing proportions of zero), the categories were collapsed for analysis.

The assumptions of normality and sphericity of the data were checked for parametric inferential statistics to be used in the analysis. If the data deviated from normality, nonparametric statistics were used. If sphericity was violated, the appropriate degrees of freedom adjustment for the ANOVA was used for interpretation of the statistics generated. The data were analysed using a two way repeated measures ANOVA (2 conditions × 3 temperature pairs). Relevant post hoc comparions were performed as appropriate. All statistical analysis were performed at $\alpha = 0.05$ and two tailed analysis were carried out.

6.7.9 Cumulative discriminability function

The d' values of adjacent stimuli for both tasks were cumulated to visualise the total discriminability across the temperature range. Durlach & Braida (1969) named the resultant plots cumulative discriminability functions (CDF). The lines of best fit through the origin were plotted using the least-squares method for the CDF of both tasks. The Weber fraction was calculated for each using the CDF. The Weber fraction, in this context, may be defined as the stimulus difference that is needed to produce a performance of $d' = 1$ as the just noticeable difference.

6.7.10 Relative judgmental variance

Equation 6.3 was used to estimate the relative variance. This is reproduced here as $\sigma^2_{MRT} / \sigma^2_{CRT} = \left(d'_{CRT} / d'_{MRT} \right)^2 - 1$, where d'_{MRT} is the discriminability between the two adjacent temperature in the MRT, d'_{CRT} is the discriminability between the two temperatures in the CRT, σ^2_{MRT} is the judgmental variance associated with the MRT, and σ^2_{CRT} is the stimulus variance associated with the CRT.

6.8 Results

6.8.1 Receiver Operating Characteristic curves

A total of 36 ROC curves were obtained. Out of these ROC curves, two curves significantly differed from the unequal-variance model at the 0.05 significance level according to the Chi-square goodness-of-fit statistic. The individual data from all subjects were jackknifed, following the approach by Dorfman & Berbaum (1986), to generate 6 additional ROC curves to summarise the results of all stimulus pairs in both tasks. The software RScore-J, by Dorfman & Berbaum (1986), was used to perform the jackknife analysis (http://perception.radiology.uiowa.edu/, accessed 18th Aug 2003) These ROCs are shown in Figure 6.3. The jackknife procedure aims to avoid the common drawbacks of conventional averaging of discriminability estimates (Macmillan & Kaplan, 1985). One of the drawbacks is obtaining a lower estimate of discriminability compared with the discriminability estimates that would be obtained from the original data if no averaging was used.

The ROC slopes for the discrimination and direct scaling methods based on the three stimulus pairs were 1.01 (S.E. = 0.09) and 1.05 (S.E. = 0.13) respectively. The slopes for both tasks did not depart systematically from unity (discrimination: $t(2) = 0.111$, $p = 0.922$; direct scaling: $t(2) = 0.385$, $p = 0.738$), therefore the Gaussian equal-variance d' was used instead of d_a.

6.8.2 Discriminability results

Figure 6.3 summarises the discriminability of the stimulus pairs within each task. Although the data were jackknifed to generate the ROC curves in Figure 6.3, the conventional averaging of the discriminability means was retained in Figure 6.4 to show the actual data for the six participants. Figure 6.3 shows that the average discriminability of the CRT task was always higher than the MRT task. Also, the discriminability of both tasks increased with an elevation of the temperatures of the stimulus pair.

The Shapiro-Wilk test showed that the discriminability data did not deviate significantly from normality. The Mauchly's test showed that the assumption of

sphericity had not been violated for the main effect of temperature pair ($\chi^2(2) = 0.786$, $p = 0.675$) and the interaction of condition × temperature pair ($\chi^2(2) = 1.947$, $p = 0.378$). The main effect of condition was not tested for sphericity violation because it has less than 2 degrees of freedom. A repeated measures ANOVA (2 tasks x 3 stimulus pairs) performed on the discrimination ability data showed a significant main effect of task ($F(1,5) = 24.98$, $p = 0.004$). There was also a significant main effect of stimulus pairs ($F(2,10) = 5.37$, $p = 0.026$). Contrasts showed that discriminability estimates for the 46-47°C stimulus pairs were not significantly different from the 45-46°C stimulus pairs with a large effect size ($F(1,5) = 2.63$, $p = 0.166$, $r = 0.59$). The contrast also showed that discriminability estimates for the 47-48°C stimulus pairs were significantly higher than the 45-46°C stimulus pairs with a large effect size ($F(1,5) = 7.529$, $p = 0.041$, $r = 0.60$). However, the interaction effect between task and stimulus pair was not significant ($F(2,10) = 0.894$, $p = 0.439$).

6.8.3 Cumulative discriminability functions

The cumulative discriminability functions (CDFs) were obtained using the jackknifed discriminability estimates. The d' values of adjacent stimuli were cumulated. The successive cumulative sensitivities provided coordinates on the y-axis for plotting the CDF. Figure 6.5 shows the CDFs for this study. The linear functions were fitted to the data using least squares method with the functions passing through the origin. There is a difference between the slopes of the two CDFs. The slope for the CRT is steeper than the MRT, indicating that the overall discriminability of the CRT was better than the MRT. Since the linear fit of these functions were adequate, it may be said that the averaged discrimination performances of the participants were in accordance with Weber's law. The Weber fractions were found to be 0.026 for the MRT task and 0.015 for the CRT task.

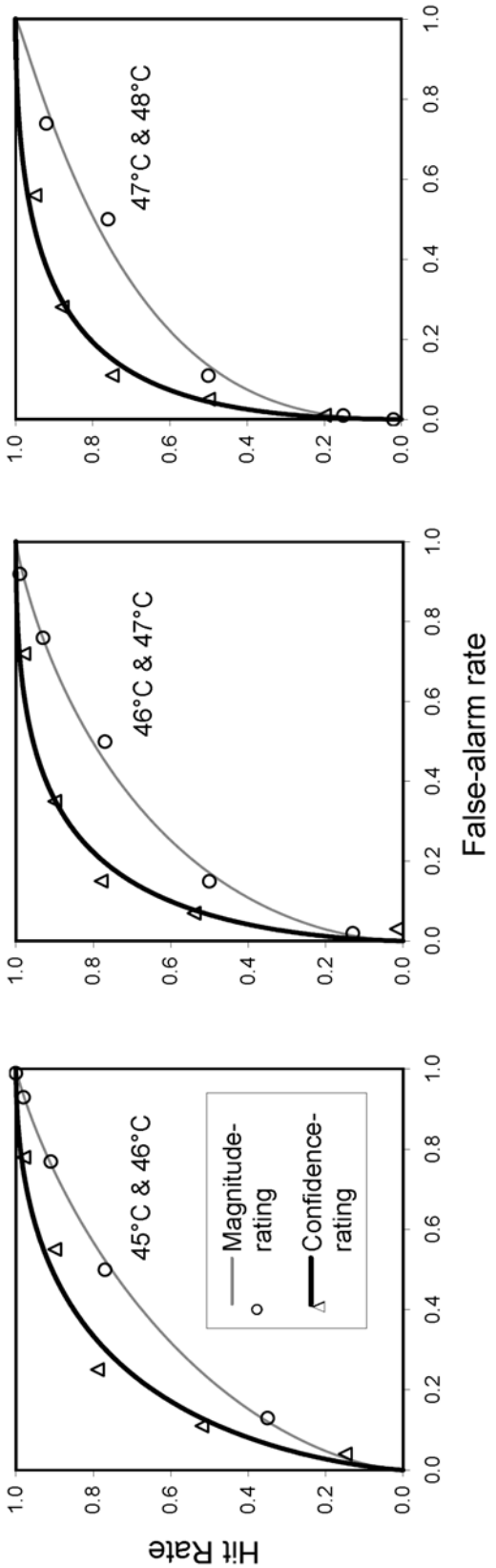

Figure 6.3. Receiver operating characteristic (ROC) curves fitted using a jackknifed procedure utilising the pooled ratings of all 6 participants. Each panel shows the ROC curves of the MRT and the CRT for each stimulus pair.

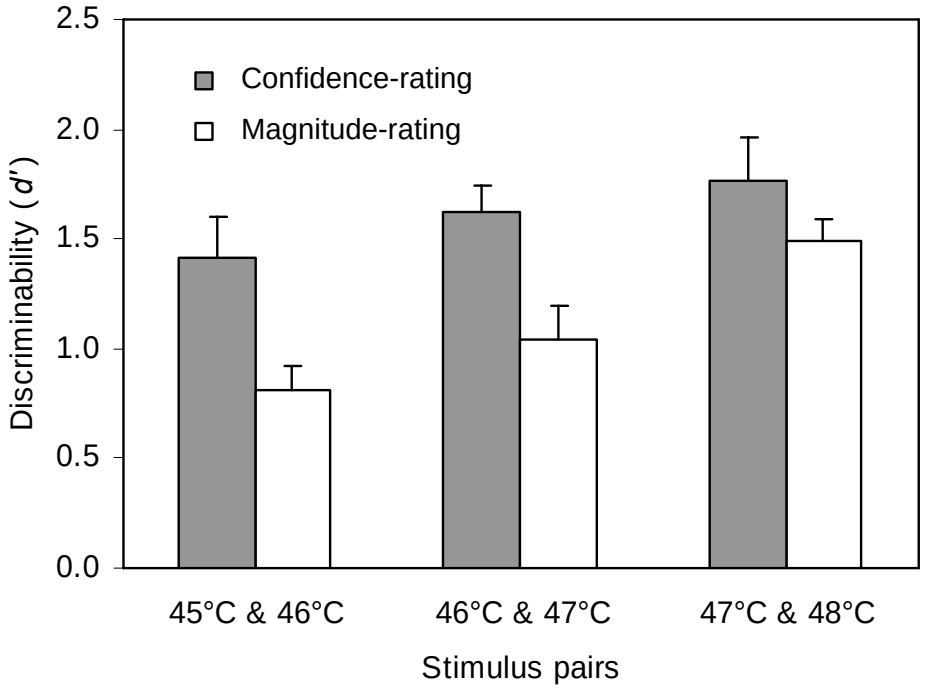

Figure 6.4. Mean discriminability, obtained through conventional averaging, of the MRT and CRT for all stimulus pairs. The error bars depict standard errors of the means.

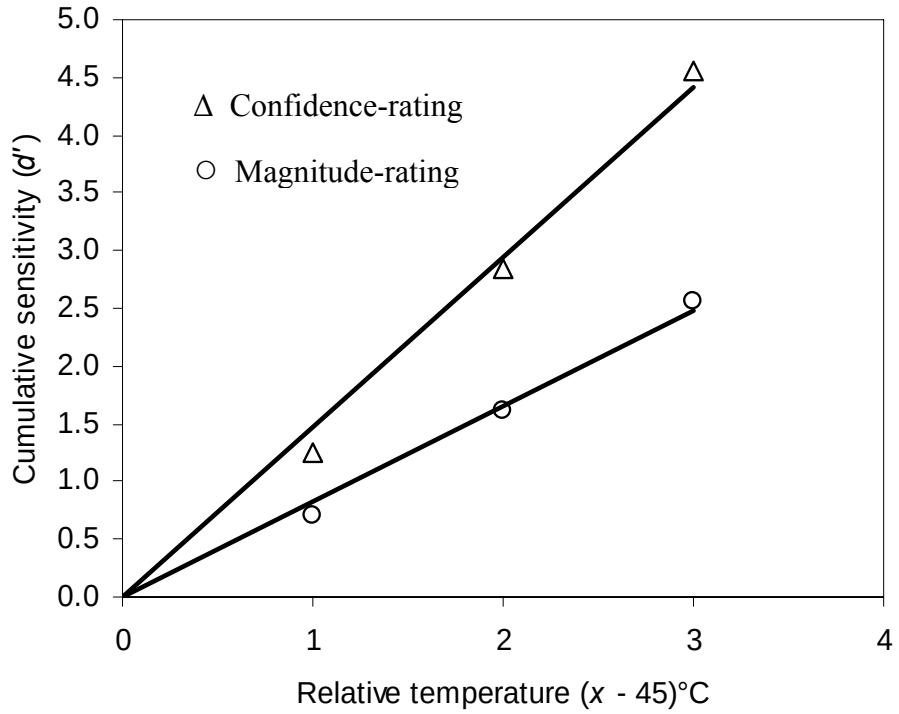

Figure 6.5. Cumulative *d′* functions for the MRT and CRT as a function of relative temperature. The jackknifed *d′* values were used to obtain the cumulative *d′* functions. The relative temperatures were obtained by subtracting 45°C from the higher temperature of each stimulus pair.

6.8.4 Relative judgmental variance

Using Equation 6.3, the additional variance in the MRT was calculated to be 2.18 times greater than the CRT. This number was calculated using the cumulated discriminability values obtained from the CDFs for the MRT and CRT.

6.9 Discussion

This study found that the MRT yielded decreased sensitivities compared to the CRT for noxious thermal stimuli. The amount of additional judgmental variance in the MRT was 2.18 times more than the CRT. These results were consistent with Durlach & Braida's (1969) predictions and previous studies (Irwin & Whitehead, 1991; Irwin et al., 1994, Rollman, 1983).

6.9.1 The contribution of judgmental variance to a poorer discriminability in MRT

This study was conducted to integrate the direct scaling and discrimination methods (MRT and CRT respectively in this study) under a common framework. The present finding of lower discriminability estimates yielded by the MRT as compared to the CRT supported the prediction that an additional component of variance may be attributed for direct scaling methods. The results suggested that outcomes yielded from discrimination methods and direct scaling methods may be related. This is demonstrated by d' obtained through MRT being lower than d' obtained through CRT for all stimuli pairs. This is predicted by Irwin & Whitehead's analytical framework. Therefore, this finding adds evidence to the assertion that discrimination methods were suitable for measuring responses from noxious stimuli. However, mine results would have to be interpreted under the framework and assumptions of Durlach & Braida's (1969) theory. Irwin et al. (1994) stated that if this same analytical framework was extended to the method of magnitude estimation, similar results could be expected.

6.9.2 Lower relative variance in this study compared to previous studies

It is interesting to note that the relative variance between the two methods found in the present study was 2.18, lower than those found by Irwin & Whitehead (1991). The

relative variance in their description (similar to the MRT) and identification task were 5.4 and 2.22 times more than the discrimination task respectively. This result is, perhaps, not unexpected, with two possible explanations. The first reason may be associated with the use of relatively lower trial numbers in this study, and the second reason may be connected with the stimulus range used for the MRT.

6.9.3 Influence of lower trial numbers on judgment variance

The introduction of lower trial numbers would inevitably increase both the variability of the responses and the likelihood of extreme proportions. This response variability may contribute considerable statistical bias to the discriminability estimates (Hautus, 1997). In Experiment D of Chapter 5, it was found that when the number of trials per intensity in an one-interval, confidence-rating task was decreased from 40 to 17 trials, the amount of variance for the discriminability estimates of the 17-trial task increased 1.74 times.

One might argue that higher trial numbers could be used to suppress the amount of variance in the discriminability estimates and it is acknowledged that using more trials should be done as much as is practically possible. There are, however, also other considerations when large trial numbers are applied, such as the onset of heat injury and hyperalgesia (Pedersen & Kehlet, 1998a, 1998b), the ethical acceptability of prolonged noxious stimulation (Charlton, 1995), and ultimately the transferability of the laboratory protocol to clinical studies. All of these factors should be carefully considered when deciding on the trial presentation numbers.

6.9.4 Influence of stimulus range on judgment variance

Another factor that may have influenced the amount of judgment variance for the MRT was the range of stimuli judged. For the CRT, the participant was only required to concentrate on the difference between the stimuli pair presented. This is in contrast to the MRT, where participants may also have compared the sensation magnitude of the presented stimulus to the context of the stimulus range. This is despite the introduction of the trial-by-trial feedback provided to participants. A similar explanation was also offered by Rollman (1979) based on adaptation-level theory (Helson, 1964). Durlach & Braida (1969) theorised that when the stimulus range was large for the scaling task, the task became more difficult for the participants. This

would lead to lowered discriminability estimates. Since discrimination tasks, in Durlach & Braida's theory, are easier in terms of the absence of the judgmental component, the performance will always be better than the direct scaling tasks. However, they predicted that for a small stimulus range, the contribution of the judgment variance in direct scaling becomes almost negligible and performance on the direct scaling task would be similar to that for the discrimination task. This prediction was generally supported by Pynn, Braida & Durlach's (1972) study on auditory intensity discrimination. This raises another possible reason for the lower additional variance observed for the MRT task in this investigation compared to other studies. There is the possibility that the temperature range for this study was relatively narrow, thus causing smaller values of the judgment variance to be found, as predicted in Durlach & Braida's theory. Nevertheless, further studies need to be conducted to confirm this conjecture.

6.9.5 CDF as a potential tool for investigating suprathreshold discriminabilities

The perception of noxious experimental stimuli has also been studied with methods obtaining point estimates of the transition between innocuousness and painfulness. (Graven-Nielsen et al., 2001). An example of point estimates used in pain research is the determination of the pain threshold using the method of limits. The effectiveness of pain relief treatments have been largely evaluated based on the lowering of this threshold. It does not illuminate the effects of pain relief treatments on the suprathreshold discriminabilities in which pain, the construct of interest, resides. This is especially important for suprathreshold discriminabilities in studies examining nociception. The same criticism could be leveled at the discriminabilities estimates obtained for individual stimulus pairs in this present study. The discriminability estimates provided information confined only to that specific stimulus pair, which reveals little about the sensitivities contained within the sensory range of interest. This problem was solved, for the purposes of this study, through the use of cumulative discriminability functions. Cumulative discriminability functions may provide additional information on the suprathreshold range of discriminabilities and the effects of intervention on them (Gracely, 2006). It may be a valuable tool for future studies investigating the description and influence of interventions on suprathreshold discriminabilities.

6.10 Conclusion

This study demonstrated that the discrimination approach (CRT) is comparable to the direct scaling approach (MRT). The bridge between the two approaches was possible through analysing the data under the theoretical framework of Durlach & Braida (1969), based on the assumption of perceptual one-dimensionality. This study's results were consistent with Durlach & Braida's prediction that an additional component of judgment variance contributed to the decreased discriminability in the direct scaling approach. It is possible to relate and compare the findings of the discrimination and direct scaling methods in studies using noxious thermal stimuli. Therefore, this framework may serve as a potentially useful tool for evaluating noxious thermal discrimination ability. This addresses specific objective 2 of this thesis (p.6) by investigating the comparability of the MRT and CRT.

For all other studies within in this thesis, the confidence-rating task will be used for obtaining the discriminability of the participants. This choice is based on: a) this chapter's verification that the CRT procedure generates less variance compared to the MRT and, b) the increased interpretational clarity of the discriminability when the MRT procedure is used (as explained in Section 4.4.2 in Chapter 4).

Chapter 7
Noxious thermal discrimination within a topical local anaesthetic state

7.1. Introduction

In Chapter 6, the comparability of the confidence-rating scale with the magnitude-rating scale was established. This chapter used the confidence-rating scale to obtain participant responses for the signal detection theory (SDT) analysis. The study within this chapter investigated the construct validity of discriminability for the specific context of analgesia induced through a topical local anaesthetic, eutectic mixture of local anaesthetics (EMLA®). In other words, this study will examine if discriminability is decreased under the local anaesthetic condition as compared to a control condition.

Although there have been many studies investigating the effects of topical local anaesthetic on sensory threshold, pain threshold and pain rating measures, few studies have looked at the effect on discrimination measures. For example researchers have investigated the depth and duration of EMLA induced analgesia (Bjerring & Arendt-Nielsen, 1990), the effect of EMLA to thermal sensory and pain thresholds (Arendt-Nielsen & Bjerring, 1988; Arendt-Nielsen & Bjerring, 1989), the effect of EMLA on capsaicin-induced pain (Fuchs, Pappagallo & Meyer, 1999) the effect of EMLA cream compared to EMLA patch (Egekvist & Bjerring, 1997), and the effect of EMLA on lithotripsy-induced pain (Tritrakarn et al., 2000). Sensory threshold, pain threshold and pain rating measures usually represent a point estimate of the transition between no sensation to sensation, between innocuousness to noxiousness or the incidental pain intensity experienced. Discrimination measures, on the other hand, usually represent the ability to differentiate between stimulus intensities.

Two studies have used SDT for analysis of the effect of local anaesthetics. One was an animal study by Lineberry & Kulics (1978) and the other, a human study by Irwin et al. (1994). Both studies used noxious electrocutaneous stimulation. Lineberry & Kulics's (1978) study examined the effects of the partial local block of peripheral

nerves surrounding the test site by lidocaine (1% and 2%) injections. Irwin et al's (1994) study examined the effects of a topical local anaesthetic on the test site. Both studies showed that SDT was a suitable model for assessing the onset of cutaneous anaesthesia. This conclusion was based on the observation that discrimination ability decreased with increasing local anaesthetic dosage, and discrimination ability increased with elapsed time of local anaesthetic removal. However, neither study compared the local anaesthetic condition with a control condition and this was a limitation. Therefore, the study within this chapter will examine the effect of a topical local anaesthetic compared with a control group.

7.2 Hypothesis

The null hypothesis for this study is:

- There is no statistically significant difference in d' change (pre- and post-EMLA) between the EMLA® condition and the control (no EMLA®) condition using noxious thermal stimuli administered on healthy individuals.

7.3 Methods

7.3.1 Study design

A within-subject, repeated measure, controlled experimental design was used.

7.3.2. Participants

A convenience sample of 10 healthy participants was recruited from staff and students of Queen Margaret University. Requests to potential volunteers for study participation were made through two methods. The first method was recruitment of students and staff by word of mouth within the Physiotherapy Subject Area, Queen Margaret University Edinburgh. The second method involved the researcher advertising the study to physiotherapy undergraduate and postgraduate students at the beginning of their classes. No participant had prior experience of the experimental protocol.

7.3.3 Inclusion and exclusion criteria

The inclusion criteria for this study were any healthy individuals studying and working at Queen Margaret University. The exclusion criteria were known congenital, idiopathic or drug-induced methaemoglobinaemia; known allergy to local anaesthetics; liver disease; receiving Class I antiarrhythmic drugs; pregnancy; medical conditions causing anaesthesia or hypoalgesia to either one of the forearms; open wounds on the stimulation and local anaesthetic application site. The researcher verbally listed the exclusion criteria to the volunteer. If the volunteer met any of the above exclusion criteria, he/she was not included into the experiment. None was excluded on this basis. All of the exclusion criteria, excepting the criteria for the presence of open wounds, were contraindications and precautions recommended by the manufacturers of EMLA for preventing potential drug interactions and aggravation of existing medical conditions (AstraZeneca, available at http://www.astrazeneca-us.com/pi/EMLA.pdf, accessed on 26th Feb 2004).

7.3.4 Ethics approval

This study was approved by the Queen Margaret University Ethics Committee. Before the experiment, all participants were briefed on the procedures of the study and their right to withdraw from the study at any time. All participants agreed to participate by signing an informed consent form in the presence of the investigator.

7.3.5 Power and sample size calculations

Based on the results of Irwin et al. (1994), the average decrease in noxious discrimination ability after topical local anaesthesia was approximately $d' = 0.80$. This was taken as the difference that this present study considered as moderate local anaesthesia. The standard deviation of d' values was approximately 0.70. This figure was obtained from the methodological study performed using 17 trials of stimuli per temperature (see Sections 5.20-5.25). The power of this study was set at 0.80 with $\alpha = 0.05$. Cohen's (1988) d for effect size was used to estimate the sample size required for this present study. An online sample size calculator was utilised for this computation (http://calculators.stat.ucla.edu/ powercalc, accessed 26 Feb 2004). The estimated sample size required was nine participants. Ten participants were recruited to anticipate any participant drop-out.

7.3.6 Equipment and testing environment

The Somedic Thermotest was used for this experiment. A description of the Somedic Thermotest was provided in Chapter 5, Section 5.3. The testing was performed in a room within Queen Margaret University described in Chapter 5, Section 5.5. The equipment set up was the same as Chapter 6, Figure 6.1.

7.3.7 Generation of local anaesthesia

Local skin anaesthesia was produced using a cream-based eutectic mixture of local anaesthetic (EMLA). The EMLA cream is an oil-in-water emulsion of lidocaine (25mg/ml) and prilocaine (25mg/ml). EMLA was chosen because it provided better local anaesthesia compared to other topical anaesthetics (Friedman, Fogelman, Nouri, Levine & Ashinoff, 1999). Various studies that used EMLA have reported no side effects associated with its use (Arendt-Nielsen & Bjerring, 1988, 1989; Fuchs et al, 1999; Mattsson et al., 1999; McDonnell & Warden-Flood, 2000).

7.3.8 Experimental conditions

There were two within-subject experimental conditions: the anaesthetised forearm versus the non-anaesthetised forearm. The stimulus set for both conditions consisted of 4 temperatures: 45°C, 46°C, 47°C and 48°C. Three stimulus pairs in total were used: 45°C and 46°C, 46°C and 47°C, and 47°C and 48°C. For the anaesthesia condition, all the stimulus pairs were presented before and after the onset of local anaesthesia in order to obtain the decrease in cutaneous thermal discriminability. The time period between the ‘before’ and ‘after’ measurements was approximately 60 minutes. For the control condition, a similar duration of 60 minutes was introduced between the ‘before’ and ‘after’ presentations to standardise the procedures for both the anaesthesia and control conditions.

There were two possible test timelines depending on the experimental condition of the first forearm that was tested. Figure 7.1 shows the sequence of events for the experiment. One timeline described the events when the anaesthetised arm was chosen to be the first testing arm (Figure 7.1A). The other timeline described the events when the control arm was chosen to be the first testing arm (see Figure 7.1B).

7.3.9 Randomisation of conditions and stimuli pairs presentation

All randomisations were performed using an online randomisation plan generator (available at http://www.randomization.com, accessed 20th Feb 2004). The following items were randomised: (a) the left or right forearm assigned to be anaesthetised for that particular participant, (b) the condition to be tested first (anaesthetised versus control forearm), (c) the administration sequence of the 3 stimulus pairs within each testing period, and (d) the trial sequences within each stimulus pair, i.e. for the 45°C and 46°C stimulus pair, whether 45°C or 46°C was given first.

7.3.10 Application of EMLA

The EMLA application procedure was monitored by a qualified podiatrist (Mr. John Veto), based at Queen Margaret University Edinburgh, experienced in the use of topical local anaesthetic.

EMLA was applied to the selected ventral surface of the forearm. It was recommended by the manufacturer that 2g of cream should be applied for every $10cm^2$ of skin surface area. It was estimated for this study that approximately 30g of EMLA was required to cover a $150cm^2$ surface area of the forearm. After application of the cream, cling-film (made from polyvinyl chloride) was wrapped over the site where EMLA was applied. The cream was left on the forearm for 60 minutes before it was removed from the skin (Figure 7.1).

It was reported by the manufacturer that following the application of 40g EMLA to $400cm^2$ of intact skin, the peak blood levels of lidocaine and prilocaine were 0.05-0.16 μg/mL and 0.02-0.10 μg/mL respectively. This is well below the toxic levels of lidocaine (> 5 μg/mL) and prilocaine (> 6 μg/mL) for causing adverse reactions such as decrease in cardiac output, total peripheral resistance and mean arterial pressure (available at www.astrazeneca-us.com/pi/EMLA.pdf, accessed 26th Feb 2004). This present study therefore used EMLA dosages that were well below recognised toxic levels.

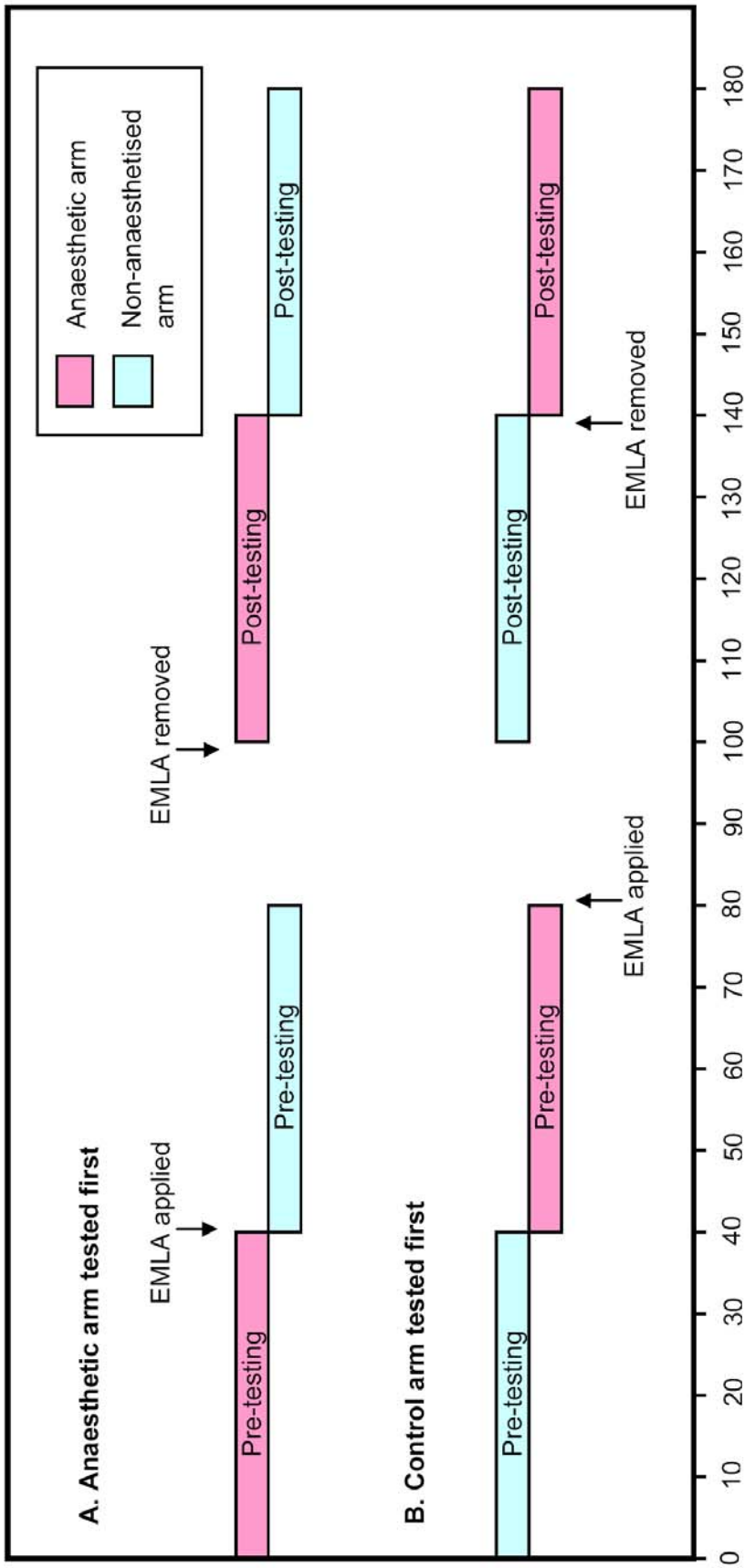

Figure 7.1. The timeline for testing of the anaesthetic and control conditions. There were 2 possible timelines: A. when the anaesthetic arm was tested first, and B. when the non-anaesthetised arm was tested first. There were 60 minutes between the pre-post testings for all conditions.

7.3.11 Psychophysical testing protocol

The testing protocol used in this study was similar to the study in Chapter 6. Twenty practice trials, similar to the actual trials, were presented at the beginning of every task for familiarisation. The one-interval rating task was used. Each trial began with the experimenter instructing the participant to place his/her forearm on the thermode (preset at the relevant testing temperature). A trial contained an observation period of 3 seconds. An automated auditory signal indicated to the participant to remove his/her forearm from the thermode. If the participant was not able to tolerate the full length of stimulus application, he/she was allowed to lift the forearm away from the thermode, although no participants did so during the study. There was an interstimulus interval (ISI) of 10 seconds before the next trial started.

As mentioned before, there were three stimulus pairs in total: 45°C and 46°C, 46°C and 47°C, and 47°C and 48°C. Each stimulus pair was presented within each test block. Within each trial only one of the two temperatures for the stimulus pair was presented. There was equal probability of presentation for either of the two temperatures. The stimulus pair presentation was randomised. The three stimulus pairs clocked a total of 102 trials per forearm per participant (34 trials for each of the three stimulus pairs). For both forearms, the total trial number presented was 204.

The participants verbally indicated their judgments to the experimenter and these were recorded. The participants' responses were made based on a confidence-rating scale with six categories (Figure 7.2). Based on the scale, the participants rated their degree of confidence on whether the stimulus presented was the higher or lower intensity of a pair of stimulus intensities. The participants were told the temperature of the administered stimulus at the end of each trial, i.e. trial-by-trial feedback was provided for both tasks. The duration of the entire study for each participant took approximately 3 hours.

1	2	3	4	5	6
I am absolutely certain the weak stimulus was presented	I am fairly certain the weak stimulus was presented	I am somewhat certain that the weak stimulus was presented	I am somewhat certain that the strong stimulus was presented	I am fairly certain the strong stimulus was presented	I am absolutely certain the strong stimulus was presented

Figure 7.2. Confidence-rating scale. This scale was presented to the participant for judgment. The participant verbally provided the number describing their degree of confidence on whether the stronger or the weaker stimulus was presented.

7.3.12 Signal detection theory analysis

The signal detection theory analysis was the same as Chapter 6, Section 6.7.8. There were altogether 108 ROC curves (9 participants × 2 conditions × 3 stimuli pairs × 2 pre-post states) generated for analysis. The data was also pooled and jackknifed using the approach by Dorfman & Berbaum (1986) to generate 4 ROC curves (2 conditions × 2 pre-post states).

The cumulative discriminability curves were also plotted using the procedure described in Chapter 6, Section 6.7.9. The cumulative discriminability curves showed the overall discriminability of the two groups' noxious thermal discrimination ability for the temperature range tested (45°C-48°C). The graphical representation of the overall discriminability allowed a visual inspection of the group discriminability results. The slope was also used to estimate the Weber fraction using the computation described in Chapter 6, Section 6.7.9.

7.3.13 Statistical analysis

The assumptions of normality and sphericity of the data were checked for parametric inferential statistics to be used in analysis. If the data deviated from normality, nonparametric statistics were used. If sphericity was violated, the appropriate degrees of freedom adjustment for the ANOVA was used for interpretation of the statistics generated. For the ANOVA analysis, the d' difference between the pre- and post-measures of the same stimuli pair and condition was used. The data were analysed using a two way repeated measures ANOVA (2 conditions × 3 discrimination differences). Relevant post hoc comparisons were performed as appropriate. All statistical analysis were performed at $\alpha = 0.05$.

7.4 Results

7.4.1 Participants

Ten healthy volunteers participated in this study. One participant (man) withdrew from the study at the mid-point with no reason provided. The data from this participant was excluded from analysis. All nine participants remaining were right-handed with a mean age 27 years (range = 22-35). Seven out of the nine participants were women.

7.4.2 Linear ROC function slopes

The slopes of the linear ROC functions for discriminability were examined for all conditions. The mean slope for the anaesthetised-pre condition was s = 0.82 (S.E. = 0.07), anaesthetised-post condition was s = 1.07 (S.E. = 0.06), control-pre condition was s = 0.94 (S.E. = 0.07) and control-post condition was s = 0.96 (S.E. = 0.08) The slopes did not deviate systematically from s = 1.0 (anaesthetic pre-: $t(2) = -2.571$, p = 0.124; anaesthetic post-: $t(2) = 1.167$, $p = 0.364$; control pre-: $t(2) = -0.857$, $p = 0.482$; control post-: $t(2) = -0.500$, $p = 0.667$). Therefore, the equal variance model index d' was used instead of the unequal variance model index d_a.

7.4.3 Discrimination ability results

Figure 7.3 shows the ROC curves obtained from jackknifing the pooled data. From visual inspection, the participants' noxious thermal discrimination ability decreased after the application of EMLA. Although the EMLA-induced d' decrease was larger than the d' decrease for the control forearm condition, the EMLA d' decrease was relatively small in magnitude (d' difference = 0.30).

Figure 7.4 shows the results of all the mean estimated d' values for the individual stimulus pairs in all conditions. It was interesting to note that in Figure 7.4, the discrimination ability for the EMLA 45-46°C post condition showed the largest decrease in d'. As the temperatures for the stimulus pairs increased, this decrease in d' disappeared with all estimated d' values hovering around $d' = 1$ for the 47-48°C temperature pairs, regardless of the anaesthetic state of the forearm. The d' values for

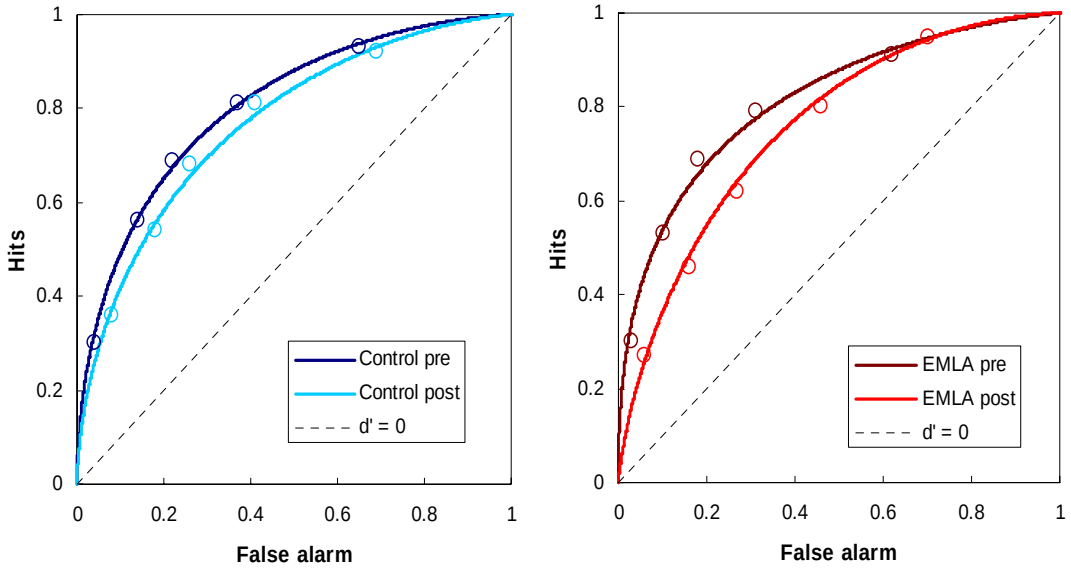

Figure 7.3. Jackknifed ROC curves representing all anaesthetised (EMLA) and control conditions. The distance between ROC curves for the anaesthetised pre-post conditions was larger than that of the control pre-post conditions. However, *d'* difference for the anaesthetised pre-post conditions was relatively small (*d'* difference = 0.3).

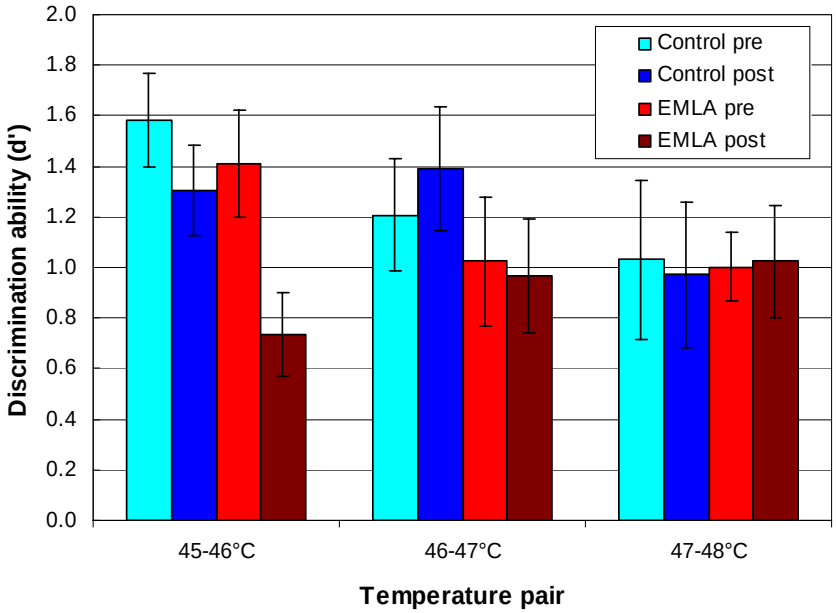

Figure 7.4. Means of estimated *d'*s for all conditions and stimuli pairs. The largest *d'* difference resides with the 45-46°C stimuli pairs where the EMLA post-test condition yielded a *d'* difference of 0.68. However, the *d'* difference was not statistically significant. All other *d'* differences were also not statistically significant. The error bars represent standard errors.

the anaesthetised conditions tested with the 46-47°C and 47-48°C were of moderate performance (*d′* range = .97 – 1.02).

Figure 7.5 was plotted to show the *d′* differences between the pre-post states of all the stimuli pairs. The results presented in Figure 7.5 are a function of Figure 7.4. This figure illustrates more clearly the differences between the pre-post states of all the conditions. The largest *d′* difference (*d′* difference = 0.68) was the 45-46°C stimulus pair for the EMLA condition. For the control conditions, the *d′* differences between the pre-post conditions were close to *d′* = 0. This meant that there were no descriptive *d′* differences for the pre-post conditions.

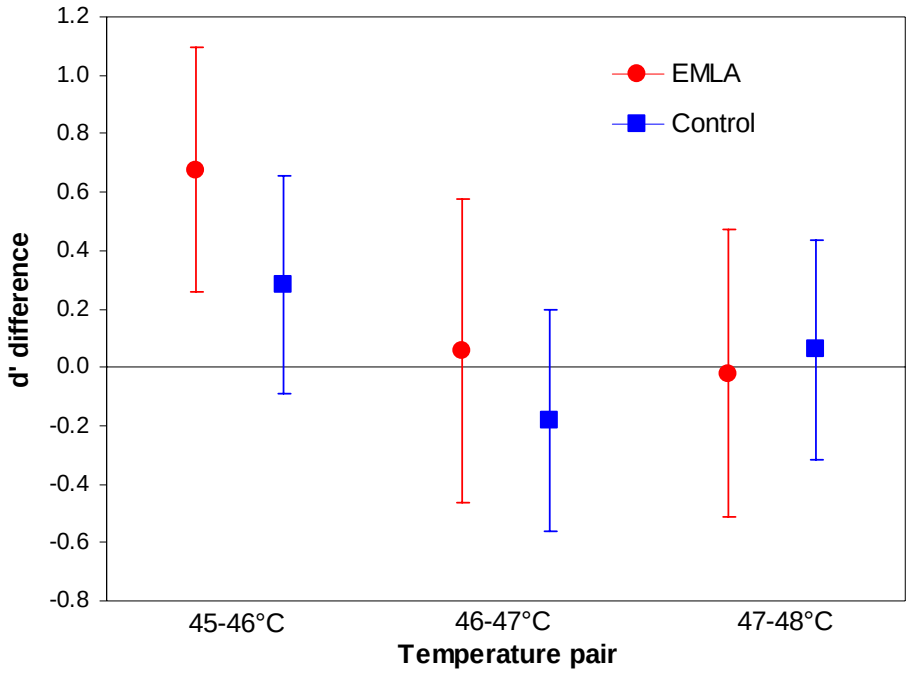

Figure 7.5. Mean *d′* pre-post differences for the conditions and stimuli pairs. The mean differences decreased with increasing temperatures for the stimuli pairs. The error bars represent 95% confidence intervals.

7.4.4 ANOVA

A two-way repeated measure ANOVA (2 sensory states × 6 pre-post d' differences) was used to examine the data. The Shapiro-Wilk test for normality showed that all variables for pre-post d' differences met the normality assumption, except for the pre-post d' difference of the 45-46°C control condition ($W(9) = 0.807$, $p = 0.025$). Mauchly's test for sphericity showed that there were no violations of the sphericity assumptions for the main and interaction effects of the dependent variables of pre-post d' differences. The ANOVA is fairly robust for mild violations of normality (Howell, 2007, p316) and the data strongly demonstrated sphericity. Therefore, the ANOVA analysis was carried out despite the deviation from normality for one of the variables.

The ANOVA showed that there were no significant main effects of the sensory states (EMLA vs. control) ($F(1,8) = 0.784$, $p = 0.402$) and pre-post d' differences ($F(2,16) = 2.368$, $p = 0.126$). The results also showed that there were no significant interactions between the sensory states × pre-post d' differences ($F(2,16) = 0.610$, $p = 0.556$).

7.4.5 Cumulative discriminability functions and Weber fractions

The d' values within each condition were cumulated to obtain the cumulative d' for the temperature range of 45°C to 48°C. Braida & Durlach (1972) named these functions cumulative discriminability functions. Figure 7.6 shows the cumulative discriminability functions for both groups. It is a useful tool that may be used for comparison of discrimination ability between studies using different temperature ranges and temperature differences in the stimuli used. Information extracted from the cumulative discriminability function for the computation of Weber's fraction may also be used as a comparison statistic (Irwin & Whitehead, 1991; Irwin et al., 1994). When the just noticeable difference is defined as $d' = 1$, the Weber fraction may be obtained using the following equation, $k = \Delta T / T$, where k is Weber's fraction which is a constant, ΔT is the just noticeable difference in intensity of the stimulus and, T is the baseline intensity (Gescheider, 1997, p3). The following Weber fractions were obtained for these conditions: EMLA pre-test ($k = .019$), EMLA post-test ($k = .025$), control pre-test ($k = .017$), and control post-test ($k = .018$).

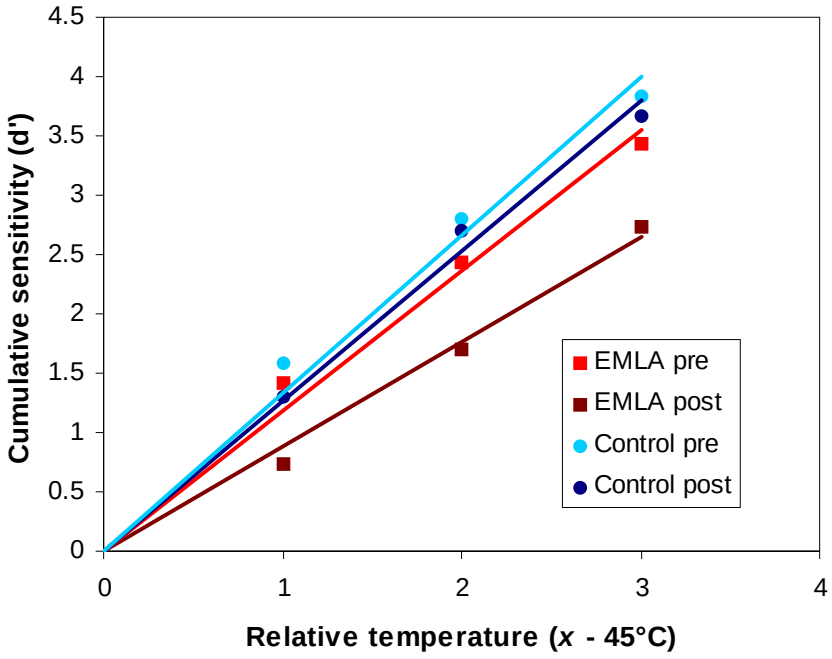

Figure 7.6. Cumulative discriminability functions for all pre- and post-test conditions. Trend lines were drawn through the data points and origin using the least squares method. The slopes of the functions represent the overall discrimination ability over the range of temperatures tested. Steeper slopes represented better discrimination abilities.

7.5 Discussion

7.5.1 Summary of findings

The descriptive findings showed that the only visible decline (d' decrease = 0.68) in noxious thermal discrimination ability was for the 45-46°C pair under the EMLA condition. The participants' discrimination ability for the other temperatures did not appear to be diminished when comparing the EMLA and control conditions. The ANOVA found that all the main effects (sensory states and d' difference) and interaction effects (sensory states × d' difference) were not statistically significant. This confirmed the descriptive findings and also verified that the apparent decrease in d' for the 45-46°C pair under the EMLA condition was not statistically significant. The Weber fraction was highest for the EMLA post-test condition (k = 0.025). This meant that it was much harder for the participants to detect an intensity change per unit measurement for the EMLA post-test condition compared to the other conditions. However, it must be noted that the relatively larger decrease in d' for the 45-46°C pair

in the EMLA post-test condition contributed much to the elevated Weber's fraction for the EMLA post-test condition. As the Weber's fraction was calculated from cumulating the adjacent *d′*s, any large decrease in *d′* from one of the stimuli pairs would depress the entire cumulative discriminability function, thus elevating the Weber fraction.

7.5.2 Noxious thermal discrimination ability

It was surprising that the analysis for the discriminability data did not yield statistically significant results between the anaesthetic and the control conditions. This was despite the observation that there was a large *d′* decrease for the 45-46°C stimuli pair in the EMLA condition. Although, in retrospect, the *d′* decrease of 0.68 might seem relevant, it must be noted that the absolute magnitude of *d′* difference when used to estimate the sample size for this present study was $d' = 0.80$. This figure was obtained from the *d′* decrease induced by EMLA in Irwin et al.'s (1994) study. Therefore, the observed *d′* difference value for the local anaesthetic effect was lower than expected. In other words, the power for this study may be too low. This finding could impact upon the determination of statistical power and sample size estimates for future studies investigating similar phenomena.

The findings of this study were contrary to Irwin et al's (1994) study. They found that discrimination ability, after the induction of local anaesthesia, increased as a function of elapsed time as the effect of the local anaesthesia slowly diminished. Also, the discrimination ability diminished for all intensities of the stimuli used in the study. This finding suggested that the *d′* changes found in their study were mainly due to local anaesthesia. This study's results also differed from Lineberry & Kulic's (1978) findings. They found that the discrimination ability of the trained rhesus monkeys diminished with local anaesthetic injection. The *d′* also decreased with increasing local anaesthetic dosage.

It was difficult to explain the results found in this present study. However, it is possible that the pharmacokinetics of EMLA might explicate the results. The following sections will briefly describe the characteristics of EMLA anaesthesia based on the parameters of the anaesthetic application. This will be followed by a proposed

explanation for the findings based on the characteristics of EMLA in relation to participants' decision-making processes for this study

7.5.3 Pharmacokinetics of EMLA

The effectiveness of EMLA analgesia may be measured by the depth of penetration and the elimination of conventional heat and pain thresholds (Arendt-Nielsen & Bjerring, 1988; Arendt-Nielsen & Bjerring, 1989;Bjerring & Arend-Nielsen, 1990). These two factors are, in turn, influenced by the duration of EMLA cream application.

Arendt-Nielsen & Bjerring (1988) found that thermal sensory and pain thresholds, measured using laser stimulation, were affected by the duration of EMLA application. Several EMLA application durations (15-120 minutes) were investigated by the researchers. Arendt-Nielsen & Bjerring's (1988) findings from the EMLA application duration of 60 minutes is relevant to this present study. It was found that the sensory threshold following 60 minutes EMLA application was significantly raised for approximately 150 minutes after EMLA was removed. The pain threshold was abolished for 90 minutes after EMLA removal. The pain threshold was raised for another 120 minutes after the initial pain abolishment.

Bjerring & Arendt-Nielsen (1990) also determined that the depth of EMLA anaesthesia is influenced by the duration of EMLA application. The anaesthesia depth was measured by the insertion of an 18-gauge intravenous needle. The participants responded when pressure was perceived. This was defined as the sensory depth. The needle was inserted further until pain was perceived. This was defined as the pain threshold depth. With increasing duration of EMLA application, there was a concomitant increase in sensory and pain thresholds. The duration of interest for this present study is the 60 minutes application. The mean sensory and pain threshold depths for anaesthesia produced by 60 minutes of EMLA application were 1.88mm and 3.11mm respectively. The maximum anaesthesia occurred 30 minutes after the EMLA cream was removed. The sensory and pain threshold remained significantly higher from the non-anaesthetic state 240 minutes after EMLA removal.

This information suggests that partial anaesthesia may have been induced in this present study. Firstly, based on Bjerring & Arendt-Nielsen's (1990) findings, it was

likely that EMLA penetrated more than 1.88mm beneath the skin surface. This depth is deeper than the most superficial C-fibre nerve endings depth, estimated to be between 20μm to 570μm (Tillman, Treede, Meyer & Campbell, 1995). Since this present study applied the EMLA in a similar way to Bjerring & Arendt-Nielsen's (1990) study, this depth of anaesthesia would most likely to have been effected as well. In Arendt-Nielsen & Bjering's study (1988), thermal sensory threshold was raised but not eliminated and pain perception was completely abolished. The reason warmth sensation persisted despite pain perception block could be due to activation of some warmth receptors located deeper in the dermis than the nociceptors. Therefore, a lower dosage of anaesthetic would have reached these deeply located warmth receptors to induce only partial anaesthesia (Arendt-Nielsen & Bjerring, 1989). Similar results were found by Arendt-Nielsen & Bjerring (1989). This indicated that the participants in this present study would likely have retained some thermal perception despite the disappearance of pain perception. This was confirmed through anecdotal reports by the participants of this present study that some thermal sensation was still perceived after the removal of EMLA. Unfortunately, the reports were not further established by evaluating participants' thermal sensory and pain thresholds. The thermal sensory and pain thresholds were not evaluated in order to shorten the test duration that participants underwent. This was to prevent the introduction of fatigue effects that may influence the participants' performance. Nevertheless, the anecdotal information was adequate to confirm that some form of thermal perception was still intact.

7.5.4 Attention to residual thermal sensations

It was probable that participants in this present study retained some thermal perception based on the pharmacokinetic studies and participant anecdotal reports. This led to a possible explanation for the discrimination ability results observed in this study. The participants' moderate discrimination ability in the EMLA conditions could be explained by the equivalence of discriminability for residual thermal sensations within the anaesthetic condition to the non-anaesthetised condition. In the control conditions, the participants may have used sensations from the noxious temperatures for the discrimination procedure. In contrast, perception of thermal sensation was slightly diminished but still present; and pain block was achieved for the EMLA conditions (Bjerring & Arendt-Nielsen, 1990). It is possible that thermal

sensation may have been used as cues for discrimination by participants. This may explain the decreased d' for the 45-46°C stimuli pair in the EMLA condition. However, this explanation still could not account satisfactorily for the lack of d' change for the higher intensity stimuli pairs in the EMLA conditions. One possibility could be that higher intensity stimuli pairs generally elicited moderate discrimination performance (approximately $d'=1$) in this study. When pain perception was blocked by EMLA, as explained earlier, the participants used the thermal sensations for the discrimination task. The resultant moderate discrimination performance based on non-noxious thermal cues might have been equivalent to the discrimination ability elicited by higher intensity stimuli pairs. This is in spite of different sensations that might have been used for the discrimination procedures. There was no mention of this suspected phenomenon in Irwin et al's (1994) study which used noxious electrocutaneous stimuli.

There are implications if participants did use innocuous sensations for the discrimination task when the aim of the experiment was to determine the extent of the local anaesthesia on pain perception. The first implication is that when discrimination measures are used to represent incompletely induced local anaesthesia, there is a possibility that discrimination might appear unaffected or the amount of sensory attenuation observed may not be an accurate reflection of the degree of anaesthesia. The second implication relates to potential methodological adjustments to account for the non-noxious discrimination phenomenon. Higher temperatures would have to be used in order to assess pain perception in an incomplete local anaesthetic state so that the stimulus would still be perceived as noxious. This also has ethical implications because the use of higher temperatures would increase the likelihood of burns that may occur through repeated stimuli application. Although this would not preclude the use of the discrimination task for environments where experimental burns or hyperalgesia were the manipulated variables. Therefore, This study's findings may not have similar implications in clinical states where perception of pain may be altered or situations involving hypersensitivity to pain perception. Further work is needed in these areas to elucidate the limitation and usefulness of discrimination as an outcome measure.

7.5.5 Effects of familiarisation and fatigue

A factor that may have influenced the consistency of the participant's responses was task fatigue. All participants underwent a three hour session. The lengthy session and the repetitive tasks might have an impact on the attentional capacity of the participants thus leading to generally decreased discriminability for the tasks performed.

On the other hand, familiarity with the task might increase participants' discriminability towards the mid point of the study. Practice trials were provided for the participants to stabilise familiarisation effects for this present study. A total of twenty trials were presented to participants before each test block: 10 trials for the higher intensity stimulus and 10 trials for the lower intensity stimulus. This was considered sufficient for participant task familiarisation without adding unnecessarily to the total trial count. Figure 7.7 displays the discrimination abilities of all participants in sequential order of stimuli pairs administered. This figure allowed a visual inspection of the individual participant's discrimination ability that is dependent on time. The discriminabilities for all participants did not show any systematic effects that might be related to fatigue or familiarisation of task and therefore can be discounted as potential explanatory factors.

7.5.6 Study limitations

The main finding for this study was the unexpected pattern of discrimination ability post-EMLA application. An explanation was proposed by which participants' utilisation of residual thermal sensation for discrimination when pain perception was abolished. There is need for further confirmation or disconfirmation of this suggestion.

In retrospect, although this study attempted to standardise the concentration of EMLA applied by volume, the degree of anaesthesia produced might have been different for individual participants. The absorption of the topical local anaesthesia into the skin depends on the composition of the skin and the duration of drug application (Arendt-Nielsen & Bjerring, 1988; Arendt-Nielsen & Bjerring, 1989; Bjerring & Arend-Nielsen, 1990). Therefore, local anaesthesia concentration could differ between participants for the same cutaneous depth. This meant that anaesthesia was only partially controlled. Perhaps a method of quantifying the degree of anaesthesia could

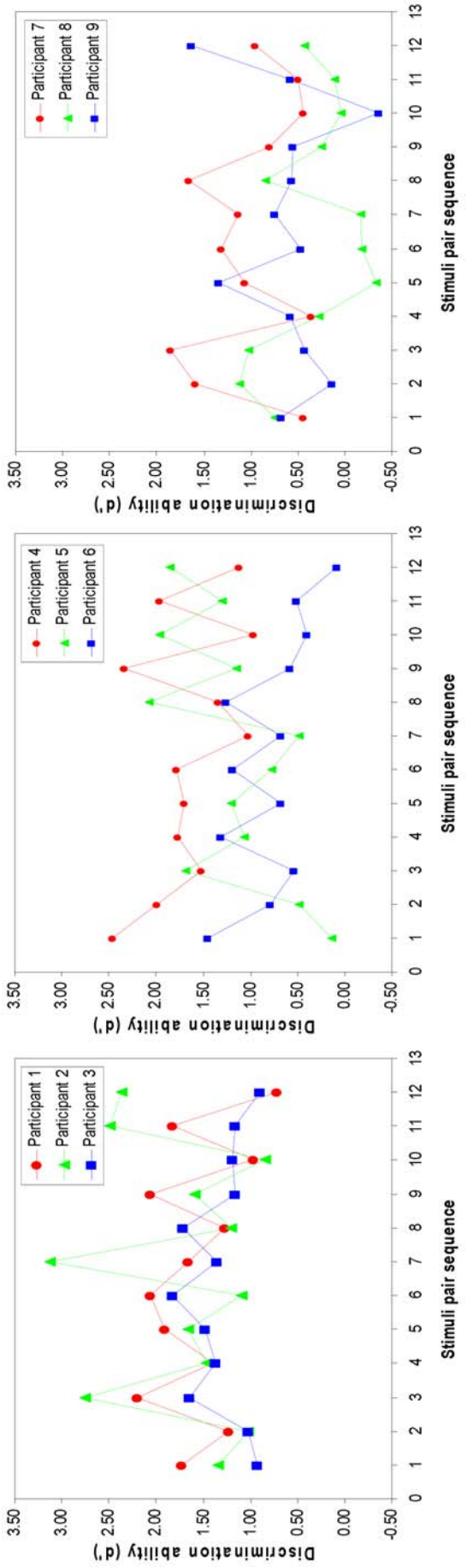

Figure 7.7. Discrimination ability of all participants arranged in the order of stimuli pair sequence administered. Visual inspection of patterns that might indicate familiarisation or fatigue effects was conducted. There was no obvious pattern that suggests familiarisation or fatigue had occurred.

be used in future studies. This would ensure that variability in discrimination ability is confined to participants' response variance rather than degree of anaesthesia variance.

7.6 Conclusion

This study found that there were no significant differences between the discrimination ability changes induced by EMLA and the control conditions. The clear but non-significant discrimination ability change for the 45-46°C stimuli between the EMLA and control conditions may warrant further investigation. The findings differed from previous studies in which alterations in discrimination ability reflected the anaesthetic state of the body region. It was proposed that the results in this study may be explained by the participants' use of residual thermal sensations instead of noxious sensation for discrimination in EMLA conditions. However, further examination of this suggestion is required. This study concluded that the utility and construct validity of discrimination ability as a measure of induced anaesthesia through a topical local anaesthetic was not established. This study addressed specific objective 3 (p.7) to examine the construct validity of discriminability as a correlate of analgesia induced using topical local anaesthetic.

Chapter 8
Noxious Thermal Discrimination in a Clinical Sample: Chronic Low Back Pain

8.1 Introduction

In the previous chapter, the construct validity of discriminability as a measure of analgesia induced by a topical local anaesthetic was not established in contrast to previous studies. In this chapter, the construct validity of discriminability as a measure of nociceptive ability in the clinical context of chronic low back pain (CLBP) sufferers compared to healthy individuals was investigated. This study also examined the correlation between discriminability with affective factors (depression and anxiety) to elucidate the amount of covariance between these variables. In addition, the correlation between potentially analgesic medication and discriminability was examined to address a methodological limitation in previous signal detection theory (SDT) studies. These investigative objectives are in direct response to the three issues outlined in Chapters 3 and 4. The three issues are: 1) the construct validity of the SDT measures when applied to interpret results of pain perception studies, 2) the potential influence of psychological factors on the SDT measures and, 3) the effect of analgesic medication on pain report by CLBP sufferers as described by SDT measures.

As discussed in Chapter 3, Section 3.1, the study within this chapter chose CLBP sufferers as the participant population for two reasons. The first reason is that most studies that have investigated painful conditions using SDT methodologies focused on CLBP sufferers (Naliboff et al., 1981; Cohen et al., 1983; Yang et al., 1985). The second reason is that these CLBP studies using SDT methodologies have produced consistent results. The discriminability of CLBP sufferers is lower compared to healthy individuals for all studies (Naliboff et al., 1981; Cohen et al., 1983; Yang et al., 1985). These two reasons meant that greater comparability of findings between previous studies (which used magnitude-rating scales to obtain participant responses) and the study within this chapter (which used a confidence-rating scale) is possible.

8.2 State-Trait Anxiety Inventory (STAI)

The STAI (Spielberger, 1983) was used in this study to gain an overview of whether the anxiety status of the participants was associated with their noxious thermal discrimination ability. The STAI is a 40 item questionnaire separated into two forms (see Appendix F). The S-Anxiety scale consists of 20 items relating to the state aspect of anxiety symptoms. The scale evaluates how participants feel 'right now' at this moment. The T-Anxiety scale consists of 20 items relating to the trait aspect of anxiety symptoms. This scale evaluates how participants generally feel. The internal consistencies of the S- and T-anxiety scales were reported to be high with respective median alpha coefficients of 0.93 and 0.90 for combined samples of students, working adults and military recruits (Spielberger, 1983, p.3). The internal consistencies of the S- and T-anxiety scales remained high for a clinical sample consisting of largely chronic pain sufferers, with median alpha coefficients of 0.94 for both scales (Novy, Nelson, Goodwin & Rowzee, 1993). The test-retest reliability coefficient at 104 days tended to be higher for the T-anxiety (Mdn_{men}= 0.73, Mdn_{women}= 0.77) than the S-anxiety scale (Mdn_{men}= 0.33, Mdn_{women}= 0.31) (Spielberger, 1983). This is expected given the transitory nature of S-anxiety. The STAI has also demonstrated good concurrent, convergent, divergent and construct validity (Spielberger, 1983).

8.3 Beck Depression Inventory (BDI)

The amount of depressive symptoms reported by participants in this study was recorded using the BDI (Beck, Steer, Ball & Ranieri, 1996). The BDI is a questionnaire consisting of 21 items that attempts to score the severity and number of depression-related symptoms experienced by an individual (Beck, Steer, Ball & Ranieri, 1996). For the present study, the BDI version 2 (BDI-II) was used instead of version 1 (see Appendix G). The substantial revisions within version 2 involved rewording of certain item options, deletion of 4 items (Body image change, Work difficulty, Weight loss and Somatic preoccupation) and the addition of 4 new items (Agitation, Worthlessness, Loss of energy and Concentration difficulty). This revision was performed so that the inventory was consistent with the guidelines of the Diagnostic and Statistical Manual of Mental Disorders IV (American Psychiatric Association, 1994). The BDI-II demonstrated good internal consistency for a psychiatric outpatient sample ($\alpha = 0.92$), the test-retest reliability of the BDI-II was

good for between test duration of one week (r = 0.93) and also demonstrated good content and construct validity (Beck et al, 1996). The inventory developers determined the optimal cut-off scores for the BDI-II through the application of receiver operating curves analysis. The following cut-off score guidelines were suggested by Beck et al, (1996) for total scores of patients diagnosed with major depression: minimal = 0-13, mild = 14-19, moderate = 20-28, and severe = 29-63. It is possible that these cut-off scores could also be used as a guide for indication of depressive symptoms in other populations, albeit with caution.

8.4 Medication Quantification Scale (MQS)

Chronic pain patients are often prescribed multiple medications for the symptoms associated with their pain. These fall under several categories. The categories include, but are not restricted to, non-steroidal anti-inflammatories (NSAIDs), aspirin/paracetamol, antidepressants, anxiolytics, muscle relaxants, benzodiazepines, barbiturates and narcotics. Quantifying medication usage poses a considerable challenge due to the potential administration of more than one medication category for a single patient. Compounding the problem of polypharmacy is the issue of equivalence of the different categories of medication with different therapeutic dosages. Two medications under different classes will have different qualitative therapeutic effects. Even for two medications within one pharmacological class that may have similar actions, some difficulty may be encountered in comparing their equivalent therapeutic dosages. This means that there is no simple way of comparing one medication with another.

The Medication Quantification Scale (MQS) was devised by Masters et al (1992) to address some of the difficulties with quantifying pain-related medication usage. The MQS is a score given to a patient based on both pharmacological classification of the medication, its associated detriment score, and the daily dosage prescribed or consumed. The MQS consists of medication classes that are, according to the scale developers, frequently prescribed for the management of pain. For each pain-related medication, a detriment weight is assigned based on the potential detrimental effects with long-term use in patients with chronic, nonmalignant pain, and a dosage level assigned based on recommended daily dosage (see Table 8.1 and 8.2). Detriment was

defined as 'the potential to produce adverse effects either acutely or with long-term use in patients with chronic, non-terminal pain' (Masters et al, 1992). These effects include, but are not limited to, addiction/physical dependence, tolerance, abuse, insomnia, and hyperalgesia. The detriment weight for each medication class is multiplied by the relative dosage level to yield a score. The scores of each individual medication are summed up to obtain a total MQS score for that patient.

Table 8.1. Pain Medication Class Detriment Weights (modified from Harden et al., 2005)

Medication Class	Detriment weight
Topical/transdermal anaesthetics, capsaicin	1.1
Antidepressants – serotonin reuptake inhibitors	1.7
Antidepressants - other	1.9
Anticonvulsants - GABAergic	1.9
Antrihypertensives	2.0
Anxiolytic – miscellaneous	2.1
Muscle relaxants – non-dependency producing	2.2
Paracetamol	2.2
Cyclooxygenase-2 inhibitors	2.3
Antidepressants – tricyclics/tetracyclics	2.3
Analgesic – miscellaneous (i.e. tramadol)	2.3
Anticonvulsants – sodium channel blockers	2.8
Sedative hypnotics	3.1
Opioids – Schedule II*	3.4
Nonsteroidal anti-inflammatories	3.4
Antipsychotics	3.6
Opioids – Schedule IV*	3.7
Opioids – Schedule III*	3.7
Muscle relaxants – dependency producing	3.8
Benzodiazepines	3.9
Steriods	4.4
Barbiturates	4.5

* The opioids classification is defined under the Unite States Controlled Substances Act, 'Schedule of controlled substances,' 21 U.S.C. 812, 03 Jan 2005 (Available from: http://www.gpoaccess.gov/index.html; accessed 08 May 2006)

Table 8.2. Relative Dosage Scores (Harden et al., 2005)

Relative dosage level	Description
1	Subtherapeutic dosage level or occasional use
2	Lower 50% of the therapeutic dose range
3	Upper 50% of the therapeutic dosage range
4	Supratherapeutic dose

8.4.1 Validity and reliability of the MQS

Masters et al. (1992) tested the concurrent validity by comparing the MQS scores of eighty-eight chronic pain patients against 12 health care professionals' clinical judgments on the desirability of the patients medication profiles. The study yielded a Spearman rank-correlation of $\rho = 0.755$ ($p < 0.0001$), a statistically significant correlation between the clinicians' desirability mean rankings and the MQS scores. In another study, Masters et al. (1992) demonstrated that the MQS is sensitive to clinical reductions in medication consumption for a chronic pain rehabilitation programme. The 12 months pre-post treatment MQS was obtained from 30 patients in a rehabilitation programme group and another group of 30 patients that were not enrolled in the programme. A 2 x 2 (groups x times) repeated measures ANOVA showed a statistically significant group by time interaction ($F(1,58) = 8.82$, $p = 0.0043$).

The preliminary inter-rater reliability of the MQS was also tested by Masters et al. (1992). Two clinicians scored 30 medication profiles according to the MQS classifications. The Spearman rank-correlation showed a statistically significant correlation between the MQS scores by the 2 clinicians ($\rho = 0.985$, $p < 0.0001$). One of the clinicians also performed another MQS scoring on the same 30 medication profile several months later. This showed a intra-rater reliability of $\rho = 1.000$ ($p < 0.0001$).

The most current version of the MQS (version 3) was used in this present study. MQS (version 3) has the latest medication classes and revised detriment weights based on the classification judgments made by 248 physicians across the United States (Harden et al., 2005). The medication classes have increased from the original 8 in MQS

version 1 (Masters et al., 1992), 10 in MQS version 2 (Kee, Steedman & Middaugh, 1998) to the current 22 classes in MQS version 3 (Harden et al., 2005). This increase in medication classes reflects the wider range of analgesic medication currently used in pain management practice (Harden et al., 2005). The overall internal consistency between ratings made by the physicians was good ($\alpha = 0.84$).

8.4.2 MQS scores calculation

The scores for the MQS were calculated according to the following criteria: a) the patient has taken the medication for the last 24 hours, and b) the medication has known analgesic properties. The MQS score for the medication taken over the last 24 hours was used to account for participants' potential 'pro re nata' (PRN) or 'as needed' usage of medication for pain relief. If the patient is prescribed a medication for daily usage, then it is likely that the medication would have been consumed within the last 24 hours. However, this may not be true for PRN medication. The MQS scores were calculated based on the formula provided by Harden et al. (2005). The MQS for each pain related medication is derived by multiplying the detriment weight and the relative daily dosage of the pharmacological class. The relative dosage level is based on the drug manufacturer recommended dosage. The total MQS score was obtained by summing all the individual medication MQS score. For example, if aspirin was consumed by a patient based on the following frequency and dosage: Six tablets per day with each tablet containing 300mg of the drug (a total of 1800mg). The detriment weight for aspirin is 3.4 and the dosage level is within the upper 50% of the therapeutic dose range, therefore a relative dose score of 3 is assigned. The MQS score for aspirin consumption is $3.4 \times 3 = 10.2$ for this patient. If other analgesic medications are consumed, then these scores are added to form a total MQS score. The online MQS calculator by drug manufacturer Grünenthal was used to calculate the MQS scores (available at http://dzf.digi-info.de/Default.aspx, accessed 20th Mar 2007). The frequency and the single dosage were entered into the calculator and a relative dosage score and MQS score was generated (See Figures 8.1A and 8.1B).

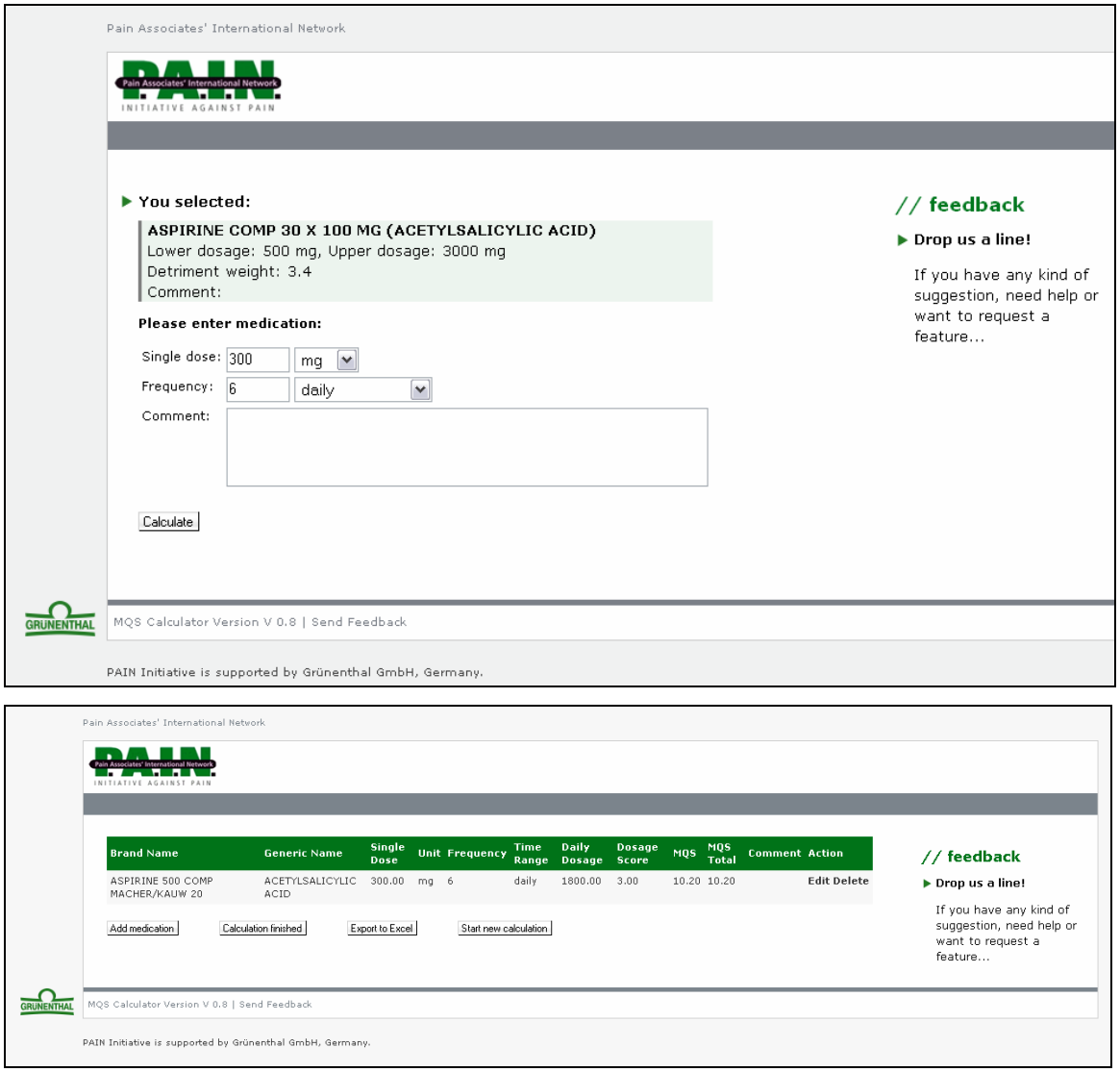

Figure 8.1. A. Online Medication Quantification Scale by Grunenthal GmbH (available at http://dzf.digi-info.de/Default.aspx). The single dose and frequency of medication were entered into the online form. B. The relative dosage score and the MQS score were generated by the online calculator.

8.5 Hypotheses

The null hypotheses for this study are:

- There is no statistically significant difference between the discriminability of CLBP sufferers compared to healthy individuals.
- There is no statistically significant correlation between the CLBP sufferers' discriminability with the level of depression (as measured by the Beck's Depression Inventory).

- There is no statistically significant correlation between the CLBP sufferers' discriminability with the level of state and trait anxiety (as measured by the State-Trait Anxiety Inventory).
- There is no statistically significant correlation between the CLBP sufferers' discriminability with the amount of potentially analgesic medication consumed in the last 24 hours (as measured by the Medication Quantification Scale).

8.6 Methods

8.6.1 Study design

The study was a pseudo-experimental study consisting of two groups: the CLBP sufferers and healthy individuals.

8.6.2 Ethics approval

The ethical approval for this study was approved by the following committees: Queen Margaret University College Research Ethics Committee and the Lothian Research Ethics Committee.

8.6.3. Inclusion and exclusion criteria

For the CLBP group, the following inclusion criteria were used: Individuals who were at least 18 years old and had a previous or ongoing history of pain in the lower back region (below L1 level) for more than 3 months. The following exclusion criteria were used for the CLBP group: participant-reported decreased cutaneous sensation on the dominant forearm (test site), participant-reported wounds or broken skin on the dominant forearm, or inability to provide informed consent.

Healthy individuals were recruited as the control group in this study. The following inclusion criteria were used for the control group: No previous or ongoing history of any painful condition in any bodily region lasting for more than 3 months. The exclusion criteria for the control group were the same as for the CLBP group with the additional exclusion criteria of currently experiencing an acute pain in any bodily region.

8.6.4 Sample size calculation

The calculations of sample size for this study were based on results obtained by Yang et al (1985). There were two groups in that study: Groups A and B consisting of CLBP sufferers and health individuals respectively. The participant numbers for each group were N_A=55 and N_B=47. The *P(A)* under the 340-390 mcal.cm^{-1}.sec^{-1} thermal pair was *P(A)* = 0.69 and *P(A)* = 0.83 for groups A and B respectively. The standard deviations for the groups' *P(A)* were $\sigma_A = 0.74$ and $\sigma_B = 0.14$. The *P(A)* values were transformed to *d'*. The transformed values were $d' = 0.707 \pm 0.4$ and $d' = 1.358 \pm 1.2$ for groups A and B respectively. Power was set at 0.8 and the α level at 0.05. An online calculator was used to estimate the sample size required for this present study based on Cohen's (1988) formula for effect size, *d*. (http://calculators.stat.ucla.edu/powercalc, accessed 26 Feb 2004). The estimated sample size per group was 31 participants. This present study consisted of two groups, therefore a total of 62 participants was required.

8.6.5 Participant recruitment strategies

Participants were recruited from several sites. Participants for the CLBP group were randomly selected from the password-protected patient database of the physiotherapy department at Western General Hospital Edinburgh. Based on the inclusion and exclusion criteria outlined, the following selection strategy was run through the computerised database: Patient has a database diagnostic code of pain in the lower back, age of patient is at least 18 years old. This generated a list of 482 potential patients that could be recruited for this study. The list was exported to the Microsoft Word ® software to personalise pre-prepared invitation letters for patients and to generate address stickers. A package consisting of the personalised pre-prepared invitation letter (see Appendix C), an information sheet of the study (see Appendix D), a reply slip consenting to being contacted by the researcher and a pre-paid addressed envelope was sent to all the identified patients. Packages were sent to all potential participants on the same day.

The patients who responded through sending back the reply slip (n = 45) were contacted by the researcher via a telephone call in the order of return. During the call, the aims and procedures were explained to the patients. The patients were given an opportunity to ask the researcher questions regarding any aspect of the study. They

were screened using the study exclusion criteria. If the patients were suitable for the study, they were requested to participate. If they verbally agreed, they were also given information regarding the location of the study and a specific appointment date and time.

Participants for the control group were recruited from staff and students of Queen Margaret University Edinburgh and a local community centre. The strategies for recruiting staff and students of Queen Margaret University Edinburgh were through word of mouth and advertising at physiotherapy classes held within the University. At the local community centre, participants were recruited through posters and publicising the study at the end of classes. People expressing an interest were given the study information sheet and consent form to make an informed judgment before deciding to participate in the study. The participants were requested to contact the researcher if they wanted to participate in the study.

8.6.6 Test site used for study

The experiment was conducted at a room located at the Western General Hospital, Edinburgh Physiotherapy Department (as reported in Experiment C of Chapter 5). Before the entire study began, accuracy and precision testing and calibration of the Somedic Thermotest were carried out at the test site (see Chapter 5, Section 5.7).

8.6.7 Demographic questionnaire

A questionnaire was designed to collect information regarding the participants' demographic details (see Appendix E). The following information was collected using the questionnaire: Age, gender, highest educational level, employment status, details of participant's general practitioner (GP), existing medical conditions, duration of low back pain, whether the participant is experiencing 'pain right now at this very moment in time', medication consumption and dosage in the last 24 hours. Information regarding the participant's GP was collected in order to inform the GP regarding the volunteer's participation in this study. The participant's medication consumption and dosage in the last 24 hours was ascertained in order to calculate analgesic usage using the Medication Quantification Scale (Section 8.4)

8.6.8 Level of education

The 2001 Census in Scotland classification for educational qualifications was used to reflect the sample of participants based within Scotland (Central Register Office for Scotland, 2003). The classification was constructed as a multi-tick question. There were a total of 7 choice items. The participants were requested to tick all the boxes that corresponded to all qualifications obtained. For this present study, a hidden scoring of the items was created in ascending order. Table 9.3 shows the classification and the hidden scoring. The highest score associated with the qualification indicated by the participant was used for analysis.

Table 8.3. Scottish Executive Census classification for educational qualifications

Which of these qualifications do you have? *Tick all the qualifications that apply*	**Hidden Scoring**
☐ 'O' Grade, Standard Grade, Intermediate 1, Intermediate 2, GCSE, CSE, Senior Certificate *or equivalent*	**1**
☐ Higher Grade, CSYS, Scottish Group Award at Higher, 'A' Level, AS Level, Advanced Senior Certificate *or equivalent*	**2**
☐ GSVQ/SVQ Level 1 or 2, SCOTVEC/National Certificate Module, BTEC First Diploma, City and Guilds Craft, RSA Diploma *or equivalent*	**3**
☐ HNC, HND, SVQ Level 4 or 5, RSA Higher Diploma *or equivalent*	**4**
☐ First Degree, Higher Degree	**5**
☐ Professional Qualifications (for example, teaching, accountancy)	**6**
☐ None of these	**0**

8.6.9 Other Questionnaires

The following questionnaires were also administered to the participants: Beck's Depression Inventory (Section 8.3) and the State-Trait Anxiety Inventory (Section 8.2).

8.6.10 Somedic Thermotest

The Somedic Thermotest was used to generate the temperatures for testing the participant's noxious thermal discrimination ability. See Chapter 5, Sections 5.3-5.4 for a description of the equipment.

8.6.11 Body region used for psychophysical testing

The dominant forearm was chosen as the region for administering the psychophysical testing. To determine the participants' upper limb dominance, they were asked to verbally indicate their preferred writing hand. The region of the forearm was the same as the studies in Chapters 6 and 7. There were two reasons for the choice of the body region tested. Firstly, an area that was fairly accessible was selected. Secondly, the tested area is similar to previous SDT studies investigating CLBP sufferers in order to increase the comparability of results.

8.6.12 Procedures for questionnaires administration

Before the psychophysical testing began, the participants were requested to complete the following questionnaires: demographics questionnaire, STAI-S, STAI-T and BDI-II. The sequence of the questionnaires was randomised before the data collection stage commenced using an online randomisation generator (available at http://www.randomization.com, accessed 14th Jun 2004). For the control group, if the participants felt that any question was irrelevant, they were asked to leave the answer blank. The participants were given as much time as they needed for answering the questionnaires. Privacy was ensured by the participants filling in the forms alone in the test room. But participants were also told that the researcher was available if help was needed.

8.6.13 Instructions for psychophysical testing

After the participants filled in the questionnaires, psychophysical testing was commenced. The participants were briefly introduced to the equipment. The protocol for testing was then explained to them. They were given the opportunity to ask the researcher any questions regarding any aspect of the psychophysical testing protocol. Standardised instruction for the actual testing was provided as follows:

> *In this experiment you will be asked to determine which one of two heat stimuli was presented to you. One stimulus is hotter than the other. Your task is to indicate whether the presented stimulus was the higher or the lower intensity and how confident you are in making that decision. There are six categories to describe your decision (Fig. 8.2 shown to the participant).*

Verbally indicate to the experimenter the category number with a description that matches most closely to your decision. After you have done this, you will be told the temperature of the heat stimulus just presented to you.

8.6.14 Psychophysical testing protocol

The testing protocol used in this study was similar to Study 2 (Chapter 6) and Study 3 (Chapter 7). Twenty practice trials, similar to the actual trials, were presented at the beginning of every task for familiarisation.

The one-interval rating task was used. Each trial began with the experimenter instructing the participant to place his/her forearm on the thermode (pre-set at the relevant testing temperature). A trial contained an observation period of 3 seconds. An automated auditory signal indicated to the participant to remove his/her forearm from the thermode. If the participant was not able to tolerate the full length of stimulus application, he/she was allowed to lift their forearm away from the thermode, although no participants did so during the study. There was an interstimulus interval (ISI) of 10 seconds before the next trial started.

The stimulus set consisted of four temperatures: 45°C, 46°C, 47°C and 48°C. These four temperatures were paired to form three stimulus pairs: 45°C and 46°C, 46°C and 47°C, 47°C and 48°C. Each stimulus pair formed a test block in which 17 trials of the lower temperature and 17 trials of the higher temperature were presented. Therefore, each test block presented a total of 34 trials. The sequence of the trials was randomised. All three test blocks clocked a total of 102 trials per participant (excluding the practice trials).

The participants verbally indicated their judgments to the experimenter and these were recorded. The participants' responses were made based on a confidence-rating response set with six categories (Figure 8.2). The participants rated their degree of confidence on whether the stimulus presented was the higher or lower intensity of a pair of stimulus intensities. The participants were told the temperature of the administered stimulus at the end of each trial, i.e. trial-by-trial feedback was provided for both tasks.

1	2	3	4	5	6
I am absolutely certain the weak stimulus was presented	I am fairly certain the weak stimulus was presented	I am somewhat certain that the weak stimulus was presented	I am somewhat certain that the strong stimulus was presented	I am fairly certain the strong stimulus was presented	I am absolutely certain the strong stimulus was presented

Figure 8.2. Confidence-rating response set. This scale was presented to the participant for judgment. The participant verbally provided the number describing their degree of confidence on whether the stronger or weaker stimulus was presented.

8.6.15 Signal detection theory analysis

The signal detection theory analysis was the same as Chapter 6, Section 6.7.8. There were altogether 372 ROC curves (62 participants × 3 stimuli pairs × 2 groups) generated for analysis. The data was pooled and jackknifed using the approach by Dorfman & Berbaum (1986) to generate 6 ROC curves (3 stimulus pairs × 2 groups).

The cumulative discriminability curves were also plotted using the procedure described in Chapter 6, Section 6.7.9. The cumulative discriminability curves showed the overall discriminability of the two groups' noxious thermal discrimination ability for the temperature range tested (45°C-48°C). The graphical representation of the overall discriminability allowed a visual inspection of the group discriminability results. The slope was also used to estimate the Weber fraction using the computation described in Chapter 6, Section 6.7.9.

8.6.16 Statistical analysis

The assumptions of normality and homogeneity of variance of the data were checked for parametric inferential statistics to be used in analysis. If the data deviated from normality or displayed heterogeneity of variance, nonparametric statistics were used. The discriminability data were analysed using a two way mixed design ANOVA (2 groups × 3 temperature pairs) to find any statistically significant main effects and interactions for the independent variables. Relevant post hoc comparisons were performed as appropriate. Correlational analysis was performed to find relationships between the demographic and psychological variables and the discrimination ability.

All statistical analysis were performed at $\alpha = 0.05$ and two tailed analysis were carried out unless otherwise stated.

8.7 Results

8.7.1 Demographics of the sample

The data collected from all participants are summarised in Table 8.4.

Table 8.4
Demographical data of participants

	Chronic LBP (n=33)	Control (n=29)
Age (mean years ± SD)	54.7 ± 13.9	57.8 ± 16.7
Level of education score [a] (median category of classification)	5.0	6.0
Duration of low back pain (years ± SD, range)	15.1 ± 10.3, 1.5-42.0	Not applicable
Gender (men/women)	13/20	10/19
MQS [b] (Median, range)	0.0, 0.0-40.0	0.0, 0.0-7.2
BDI [c] score (mean ± SD)	10.9 ± 8.2	3.9 ± 3.5
STAI-S [d] score (mean ± SD)	35.2 ± 9.3	28.3 ± 6.9
STAI-T [e] score (mean ± SD)	40.6 ± 9.9	32.9 ± 10.4

[a]The level of education was determined by Census for Scotland classification (Registry Office for Scotland, 2003). [b]MQS = Medication Quantification Scale, [c]BDI = Beck's Depression Inventory, [d]STAI-S = State-Trait Anxiety Inventory – State and, [e]STAI-T = State-Trait Anxiety Inventory – Trait.

The demographical variables were examined to see if there were any differences between the two study groups. Separate independent t-tests were used to analyse age and STAI-T because the data met parametric assumptions. The level of education, MQS, BDI and STAI-S were analysed using the Mann-Whitney test because parametric assumptions were not met by the data.

The following variables showed a significant difference between the chronic pain and control group: BDI ($U = 193.00$, $p = 0.001$), STAI-S ($U = 243.00$, $p = 0.001$) and STAI-T ($t(60) = 2.977$, $p = 0.004$). The following variables did not show a significant

difference between the two groups: age ($t(60) = .797$, $p = 0.428$), level of education ($U = 438$, $p = 0.549$), MQS ($U = 459.50$, $p = 0.764$).

8.7.2 Descriptive statistics for discrimination ability data

Figure 8.3 shows the mean discrimination ability of the three temperature pairs by both groups. In general, the d' values of both groups indicate below moderate performance (i.e. $d' < 1$). However, the CLBP group had higher d' values for all three temperature pairs than the control group based on the conventional averaging results.

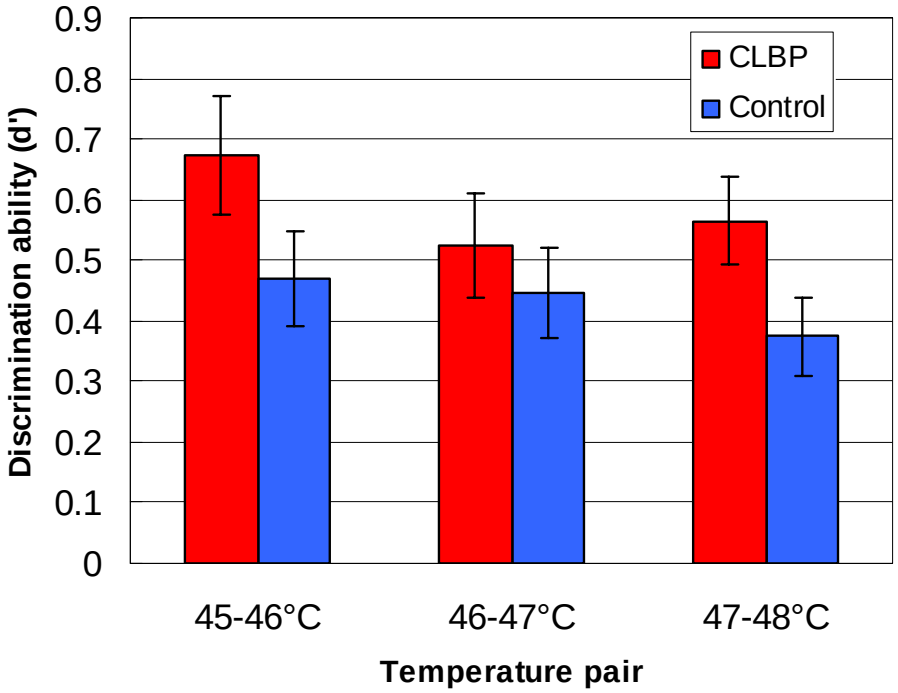

Figure 8.3. The mean discrimination ability (d') of chronic low back pain sufferers and healthy individuals in the control group for all three temperature pairs. The error bars indicate standard error.

8.7.3 Linear ROC function slopes

The slopes of the linear ROC functions for discriminability were examined for both groups. The mean slopes were $s = 1.05$ (S.E.= 0.06) and $s = 1.11$ (S.E.= 0.03) for the CLBP sufferers and healthy individuals groups respectively. The slopes do not deviate systematically from $s = 1.0$ (CLBP: $t(2) = 0.833$, $p = 0.492$; Healthy: $t(2) = 3.667$, $p = 0.067$). Therefore, the equal variance model index d' was used instead of the unequal variance model index d_a.

8.7.4 ROC curves for temperature pairs

The data for both groups were jackknifed as outlined in Chapter 6 using the approach by Dorfman & Berbaum (1986). This generated 6 ROC curves summarising the results of all the temperature pairs for both groups (Figure 8.4). The ROC curves generated generally agree with the values obtained with conventional averaging as shown in Figure 8.3. The only difference observed was for the 46-47°C temperature pair results. The jackknifed ROC curves showed that the CLBP group obtained a lower d' value compared to the healthy control group. However, the conventional averaging method yielded opposite results. This did not pose a major problem as the d' differences for the two approaches were small. The d' difference between the groups using conventional averaging and the jackknife procedure were $d' = 0.08$ and $d' = 0.02$ respectively. Practically speaking, these small d' differences were negligible.

8.7.5 Cumulative discriminability functions

The d' values for the three temperature pairs were cumulated to obtain the cumulative d' for the temperature range of 45°C to 48°C. Figure 8.5 shows the cumulative discriminability functions for both groups. Another method of comparison between studies, with information extracted from the cumulative discriminability function, is the computation of Weber's fraction (Irwin & Whitehead, 1991; Irwin et al., 1994). When the just noticeable difference is defined as $d' = 1$, the Weber fraction may be obtained using the following equation, $k = \Delta T / T$, where k is the Weber's fraction which is a constant, ΔT is the just noticeable difference in intensity of the stimulus and, T is the baseline intensity (Gescheider, 1997, p3). The Weber fraction for the CLBP and healthy control groups were 0.037 and 0.050 respectively. This meant that the CLBP pain group was able to detect a smaller intensity change per unit measurement for the stimulus compared to the healthy control group.

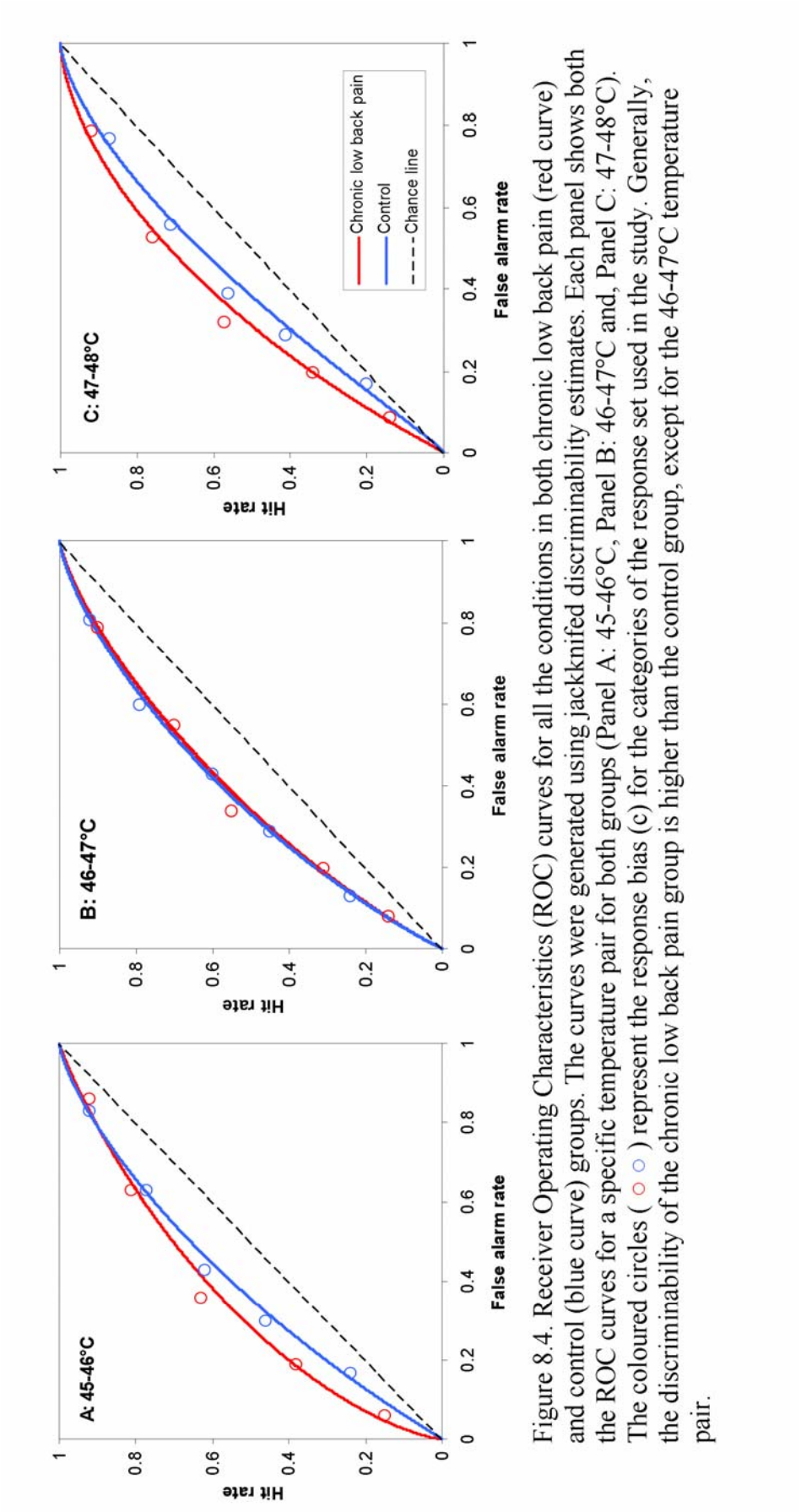

Figure 8.4. Receiver Operating Characteristics (ROC) curves for all the conditions in both chronic low back pain (red curve) and control (blue curve) groups. The curves were generated using jackknifed discriminability estimates. Each panel shows both the ROC curves for a specific temperature pair for both groups (Panel A: 45-46°C, Panel B: 46-47°C and, Panel C: 47-48°C). The coloured circles (○ ○) represent the response bias (c) for the categories of the response set used in the study. Generally, the discriminability of the chronic low back pain group is higher than the control group, except for the 46-47°C temperature pair.

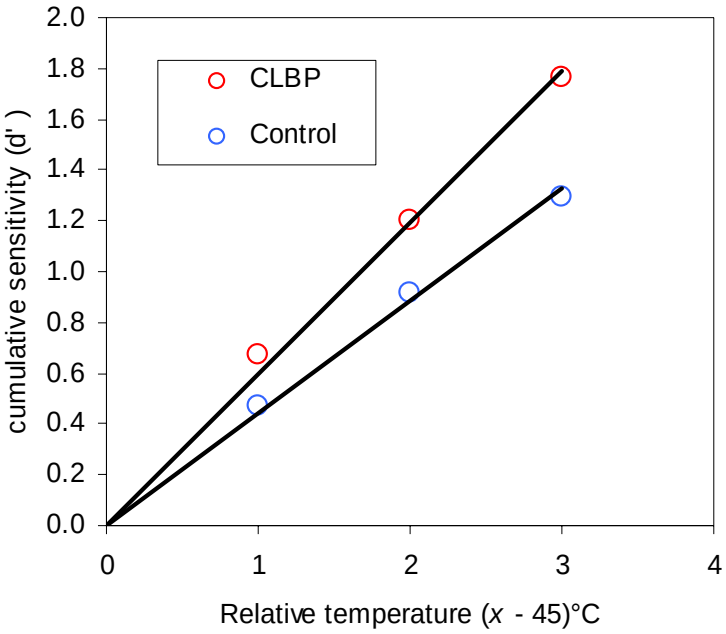

Figure 8.5. Cumulative discriminability functions for the chronic low back pain (CLBP) group and healthy control group. The function for the chronic pain group is steeper showing that the overall discriminability for the temperature range tested is higher in the chronic pain group.

8.7.6 ANOVA

Violations of normality were examined for both between and within subjects factors for the discriminability variable. The discriminability variables were also examined for violations of the assumptions of homogeneity of variance and sphericity for the between subjects factors and within subjects factors of the mixed design ANOVA respectively.

The Shapiro-Wilk test showed that the discriminability data did not deviate significantly from normality, except for the 45-46°C discriminability data by the CLBP group ($W(33) = 0.919$, $p = 0.017$). Further examination of the 45-46°C discriminability data using histograms and box plots showed that the violation of normality was contributed to by 2 outliers, participants 3 and 21 (Figure 8.6). These participants had d' values of more than 2.0 for the 45-46°C stimulus pair which were much higher than the range of the other data points. It is known that the Shapiro-Wilk test is affected by relatively large samples in which small deviations from normality yield significant results (Field, 2005, p744). These outlier data points were excluded to explore their effect on the normality of the distribution. An exploratory Shapiro-

Wilk test conducted on the remaining data showed that it did not deviate significantly from normality ($W(31) = 0.959$, $p = 0.268$).

The results for the Levene's test showed that the variances of data for all the discriminability variables were homogeneous (45-46°C discriminability: $F(1,60) = 0.701$, $p = 0.406$; 46-47°C discriminability: $F(1,60) = 0.591$, $p = 0.445$; and 47-48°C discriminability: $F(1,60) = 1.028$, $p = 0.315$). Also, Mauchly's test showed that the assumption of sphericity has not been violated for the discriminability data ($\chi^2(2) = 2.596$, $p = 0.273$).

The ANOVA is fairly robust for mild violations of normality (Howell, 2007, p.316) and the discriminability data strongly demonstrated homoscedasticity. Based on this, a judgment was made to include all data points for an ANOVA to be carried out despite the deviation from normality for one of the discriminability variables.

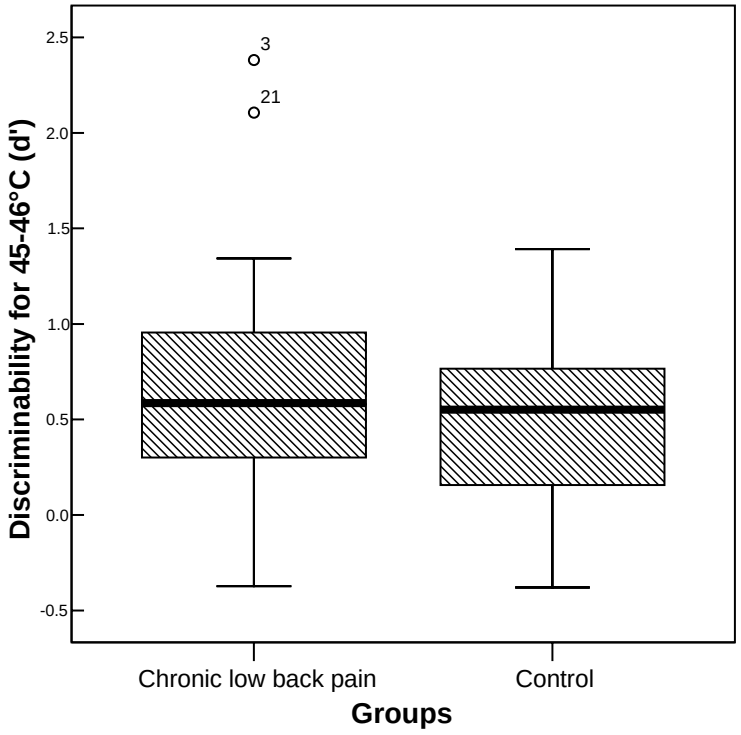

Figure 8.6. Boxplots for the 45-46°C discriminability data. The dark line within the box indicates the median of the values within the group. The error bars enclose the 95th percentile region of the data. Participants 3 and 21 are shown to be outliers for the chronic low back pain group.

The two-way mixed ANOVA showed that the discrimination ability was not significantly affected by the temperature pair administered ($F(2,120) = 1.029$, $p = 0.360$). Also, the interaction effect between temperature pair and group was not statistically significant ($F(2,120) = 0.396$, $p = 0.396$). However, the results showed that the discriminabilities between the chronic pain and control groups were statistically significant with a small to moderate effect size ($F(1,60) = 4.828$, $p = 0.032$, $r = 0.27$).

8.7.7 Correlational analysis of variables and mean overall discriminability

The associations between MQS, duration of low back pain, BDI, STAI-S and STAI-T with discriminability for the CLBP sufferers were analysed. The *d'*s were averaged across the three stimuli pairs yielding an overall discriminability for each participant in the correlational analysis. The correlational analysis was performed between the demographical variables and the average overall discriminability using the Kendall's τ correlation. There were no statistically significant correlations between average overall discriminability and the following variables: MQS ($\tau = -0.290$, $p = 0.828$), BDI ($\tau = 0.066$, $p = 0.597$), STAI-S ($\tau = 0.148$, $p = 0.597$) and STAI-T ($\tau = 0.159$, $p = 0.198$). However, the duration of low back pain was significantly correlated with the average discriminability ($\tau = 0.194$, $p = 0.039$).

8.7.8 Discrimination ability and clinical pain status

It is possible that the chronic pain patients who were experiencing pain during testing may perform differently on the task as compared with patients with no pain. One reason for this difference might be that the pain experienced by participants during testing may alter the patients' sensory, perceptual or cognitive abilities (Apkarian, Stea & Bolanowski, 1994; Bolanowski, Gescheider, Fontana, Niemiec & Tromblay 2001). This may in turn be reflected in the discriminability scores for the discrimination task.

A subgroup analysis was performed for the chronic pain group comparing the discrimination ability between the patients that were experiencing clinical pain during the task (n = 18) and those that did not (n = 15). A one way repeated measures ANOVA with the current pain of the patient designated as the between subject factor (pain status × temperature pair). A point-biserial correlation was also performed to

examine if there was an association between the CLBP sufferers's current pain status and medication consumption (MQS).

Figure 8.7 shows the mean *d'* for the 3 temperature pairs between the CLBP sufferers who experienced pain during the task and those that did not. The ANOVA results showed that there were no statistically significant differences for the discriminability of the patients that did and those that did not experience clinical pain ($F(1,31) = 0.285$, $p = 0.597$). There were no statistically significant findings for the main effects of temperature pair ($F(2,62) = 0.810$, $p = 0.449$) and the interaction effect of pain status × temperature pair ($F(2,62) = 0.108$, $p = 0.898$). The correlation between the CLBP sufferers's current pain status and medication consumption was not statistically significant ($r_{pb} = -0.04$, $p = 0.814$). However, as this analysis was not planned, caution should be taken in the interpretation of this result.

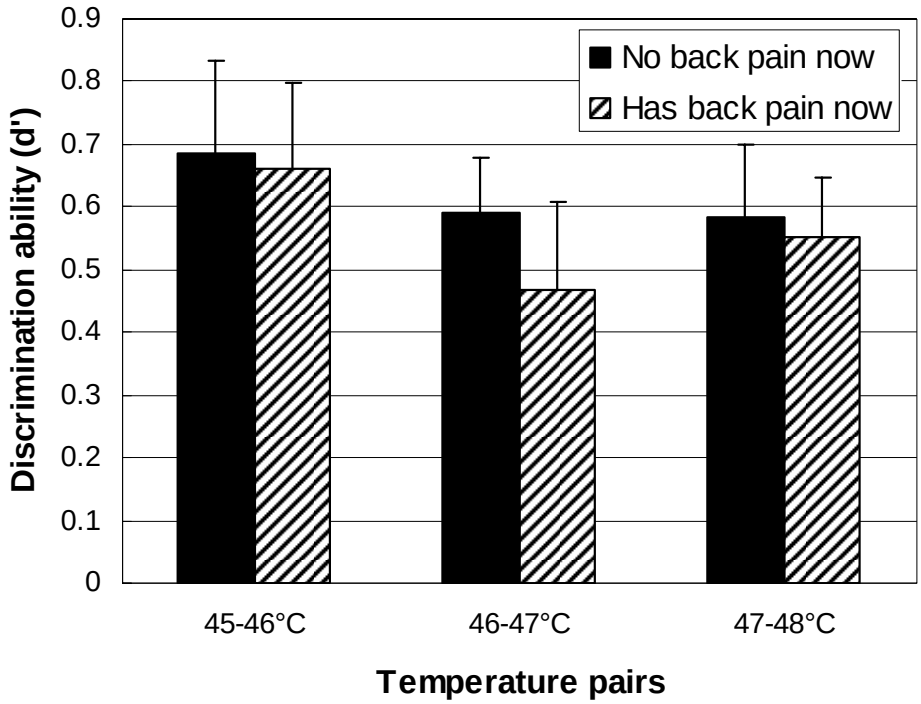

Figure 8.7. Discrimination ability of those chronic pain sufferers experiencing pain (n = 18) and not experiencing pain (n = 15) during the testing. There was no significant difference in the discrimination ability between these two subgroups of chronic pain sufferers. The error bars represent standard error.

8.8 Discussion

8.8.1 Summary of the findings

This study found that CLBP sufferers have significantly elevated discrimination ability to noxious thermal stimuli as compared to healthy individuals. The noxious thermal discrimination ability of the participants were not correlated with the amount of depressive symptoms, state-anxiety characteristics, trait-anxiety characteristics and the amount of potentially analgesic medication consumed in the last 24 hours. However, discrimination ability of the participants was positively correlated to the duration of low back pain experienced by the chronic pain sufferers.

8.8.2 Noxious thermal discrimination ability

The noxious thermal discrimination result found in this study was different to that found in previous studies investigating similar phenomena. These studies consistently demonstrated that when noxious radiant thermal stimuli were applied, the chronic pain sufferers tended to have poorer discrimination ability compared to the control group (Cohen et al, 1983; Naliboff et al, 1981; Yang et al, 1985). Several methodological variations between the methods of this study and previous research could have accounted for the different results obtained. Some of these variations were the temperature differential used for discrimination, the type of thermal stimuli administered and the rating response set used for judgment. These variations will be discussed in this section.

The temperature differential for each temperature pair in this study was set at 1°C. This was determined so as to reduce the probability of a ceiling effect for the discrimination ability (for example $d' > 2.1$ for participants 3 and 21). Similar studies have used temperature differentials of between 2.5°C (Yang et al, 1985) to 3.0°C (Cohen et al, 1983; Naliboff et al, 1981). Even though the temperature differentials between studies were slightly different, this did not provide an adequate explanation for the differences in the discrimination ability. It is possible to directly compare d' values in this study to previous research. The cumulative d' value for the two groups in this study and previous studies are presented in Figure 8.8. The d' values for this present study were taken directly from the cumulative discriminability function

(Figure 8.5). The assumption that *d′* values may be summated is adopted in this analysis (see Section 6.4, pp.109-111).

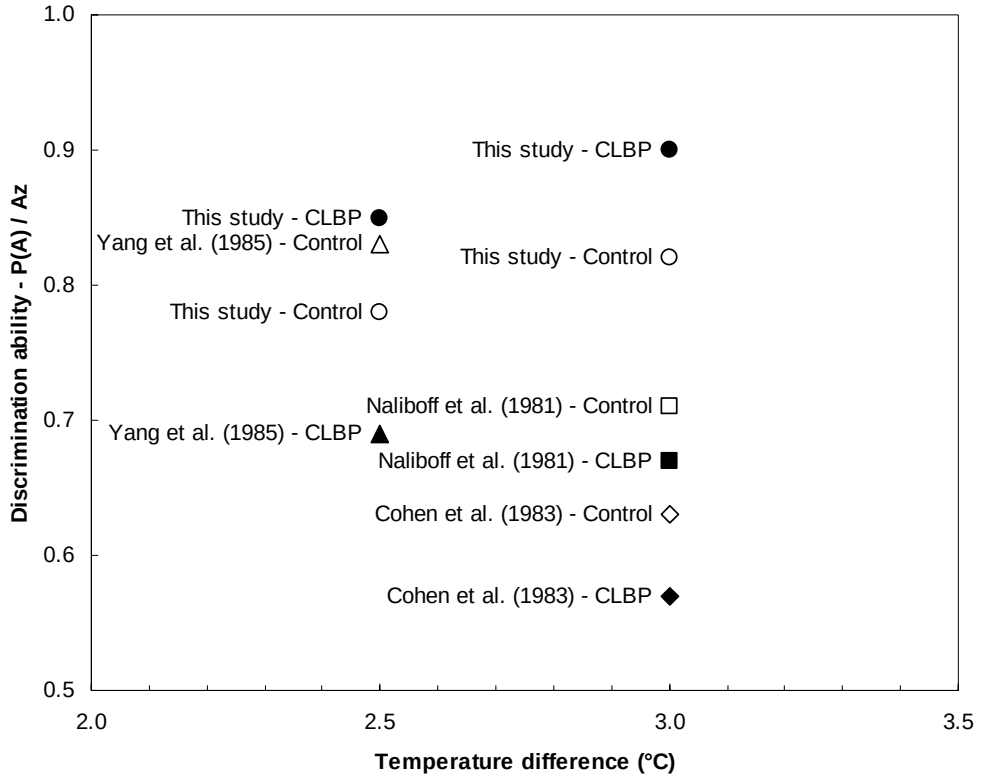

Figure 8.8. Comparison of discrimination ability between chronic low back pain sufferers and healthy individuals for this present study and previous studies. Discriminability values of 2.5°C and 3.0°C were obtained from the cumulative discriminability function of Figure 8.5. The *d′* values for this present study were transformed to A_z for comparison with the *P(A)* values in previous studies. It is shown here that the equivalent cumulative *d′* for this present study is higher compared to the other studies.

Figure 8.8 shows that even when temperature differentials have been accounted for, the CLBP groups still demonstrated higher *d′* values compared to the control group. It is interesting to note that the transformed *d′* to A_z values for all conditions in this chapter's study were higher compared to the *P(A)* values obtained by Cohen et al (1983), Naliboff et al (1981) and Yang et al (1991), for both the CLBP and control groups. A_z and *P(A)* are both 'area under the curve' indices. This allows a rough comparison of both indices. This meant that the participants in this thesis' study were

generally slightly better in discriminating between the temperatures as compared to the other studies.

There is a variation in the method of noxious thermal stimulation delivery between this thesis' study and previous studies. This study used contact thermal stimuli for the induction of nociception whereas other studies had used infrared thermal stimuli. The type of thermal stimuli administered to induce nociception was unlikely to have accounted for the increased d' values observed for the chronic pain group. Although a contact thermal stimulus may provide more tactile cues for the participants for judgment, previous studies have generally found that tactile or other innocuous stimuli generally did not affect or decrease the discriminability of the noxious stimuli (Bini, Cruccu, Hagbarth, Schady & Torebjork, 1984; Nahra & Plaghki, 2005). This might weaken the suggestion that the differences in results between this study and previous studies were due to the variations in noxious stimulation delivery.

The rating scale used in this study was a variant of the rating scale used by previous studies. The participants' confidence-rating on intensity judgment was obtained in this study. Previous studies used the category rating scale of subjective intensities. Although these two rating scales may appear to be different, the outcomes obtained may be analogous. In Chapter 6 of this thesis, it was shown that these two rating procedures produce comparable results when Braida & Durlach's (1972) judgment theory was used as a common analytical framework. The common analytical framework allows researchers to compare with caution the results of studies using these two variant procedures. Therefore, it is unlikely the type of rating scale could have accounted for the differences in discrimination for both groups in this study compared to past studies.

In conclusion, the methodological variations adopted for this thesis' study is essentially an improvement on previous SDT study methodologies. It is unlikely that the methodological variations could have accounted for the differences in discrimination as compared to previous similar studies.

8.8.3 Attentional mechanisms as a potential explanatory model

One potential model that may explain the observed findings of higher discrimination ability for the CLBP sufferers compared to the healthy individuals is the 'hypervigilance' model. This concept was first applied to pain perception by Richard Chapman (Chapman 1978, 1986). The concept suggested that individuals who use somatosensory stimuli as indications of danger were more likely to scan their body for threatening sensations. Since the individual with a predilection for hypervigilance to pain would be more alert to bodily sensation, this would lead to the prediction that these individuals would display lowered pain thresholds. Hypervigilance could therefore be viewed as a process associated with attention.

A related model named the 'cognitive-affective model of the interruptive function of pain', forwarded by Eccleston & Crombez (1999), built on Chapman's ideas of hypervigilance and models of vigilance in experimental psychology. Eccleston & Crombez (1999) proposed that pain interrupts the attention and behaviour of the individual and urges the person to escape from the pain. This interruptive nature of pain is a dynamic process which involves the interaction between the characteristics of pain and the context in which the pain resides. Three pain characteristics that would amplify the interruption of attention by pain were proposed: the intensity of pain, the unpredictability of the pain, and the threat value of the pain (Crombez, 2006). In other words, the more intense, unpredictable and threatening the painful stimulus, the more attentional interruption would be present (Crombez, Baeyens & Eelen, 1994; Crombez, Eccleston, Baeyens & Eelen 1998a, 1998b). Most studies concerned with elucidation of this model have focussed on the attention interruptive effects of a painful stimulus, either experimentally-administered or clinically-experienced, on the performance of another non-noxious attention demanding task. Generally, the performance of participants on the attention demanding task is diminished based on the model's predictions of attentional interruption (Crombez, Eccleston, Baeyens & Eelen, 1996, 1997, 1998b, 1998b). However, it is unknown how chronic pain sufferers would perform on a noxious attention demanding task, since no studies have investigated this scenario based on the hypervigilance model.

In this present study, the situation might be slightly more complex. The study context consisted of two types of noxious stimuli: the experimental noxious stimulus and the

clinical pain experienced by patients. There are two possible explanations for the chronic pain sufferers' higher d' score. The first reason is associated with factors relating primarily to the pain and its environment (Eccleston & Crombez, 1999). It is possible that certain dimensions of the experimental noxious stimulus used in mine study made it more attentionally engaging than the experienced clinical pain. For example, the perceived intensity of the noxious thermal stimulus could have been judged to be more noxious or unfamiliar compared to the clinically experienced pain. This could have made the noxious thermal stimulus more attention-demanding. The second reason is related to factors concerned with the processing of the nociceptive signals. A possible example would be functional changes that may alter the perceptual or decision-making ability of the chronic pain sufferer. For example, it has been shown that there are changes in brain morphology of chronic pain sufferers (Baliki et al, 2006) and that they demonstrate poorer ability in their emotional decision-making processes (Apkarian et al., 2004). If it is indeed assumed that pain is attention-interruptive, the increased d' could be due to the novelty of the noxious experimental stimuli or to some cognitive-affective aspect of the participants' perceptual abilities being altered in some way by the clinical pain. Unfortunately, this study was not designed to test this prediction. The exploratory analysis of the two subgroups within the chronic pain group (currently experiencing pain and no pain experienced) provided some evidence to suggest that the immediate context of the sufferer's pain may be excluded as an influence on the discrimination ability. However, caution should be taken when extrapolating this finding due to the exploratory nature of the analysis, the potential lack of power for this analysis as well as the crude two-category classification of the current clinical pain state of the sufferers. Future investigations testing a similar hypothesis could broaden the categorisation of current pain state through the use of pain intensity reporting (for e.g. visual analogue scale or magnitude estimation).

8.8.4 Correlation between anxiety and noxious discrimination ability

The STAI-S scores of the CLBP group and healthy control group fell within the 50th-69th percentile and 24th-35th percentile, respectively, when compared to a normative reference of working adults (Spielberger, 1983). The STAI-T scores of the CLBP group and healthy control group fell within the 69th-84th percentile and 47th-59th percentile, respectively (Spielberger, 1983).This present study found that the state and

trait anxiety scores were not significantly associated with the discrimination ability of the participants. Other studies were equivocal regarding the effect of state anxiety on discrimination ability. Some studies have found that state anxiety decreased the participants' discrimination ability to noxious stimuli (Malow, 1981; Malow, West & Sutker, 1989), and one study found that state anxiety did not have a significant influence on noxious discrimination ability (Dougher, 1979). This study provided some evidence to suggest that state anxiety and trait anxiety were not associated with noxious discrimination ability.

Arntz, Dreesen & Merkelbach (1991) stated that anxiety may exert a modulating effect on pain perception through attentional mechanisms. This is supported by Malow's (1981) findings which showed that anxiety-inducing threats may have diverted attentional focus away from the noxious experimental stimuli thereby resulting in poorer performance on the experimental task. In order for the anxiety-inducing threats to be effective, they were usually designed to be noxious. For example, Malow (1981) and Crombez et al. (1996) used painful electrical stimuli as the anxiety-inducing threat. In this study, however, there was no immediate anxiety-inducing threat as implied by the non-significant correlation between STAI-S scores and discriminability. The only potentially anxiety-inducing threat that may be present is the noxious thermal stimulus used for assessing participants' noxious discrimination ability. If the noxious thermal stimulus was indeed anxiety-inducing, it was not reflected in higher STAI-S scores. It may be that the threat-inducing characteristics of the noxious thermal stimuli appeared only during the testing, whereas the questionnaire was administered before testing commenced. Therefore, if the noxious stimuli were anxiety-inducing, this information was not captured.

8.8.5 Correlation between depressive symptoms and noxious discrimination ability

The depressive symptoms, as measured by the BDI-II, experienced by the chronic pain group and healthy controls were classified as minimal severity (Beck, Steer & Brown, 1996). Previous studies have found that discrimination ability to noxious stimuli in participants with depression is either lowered or not significantly different compared to healthy individuals. These studies included participants that did not experience chronic pain (Davis et al., 1979; Dworkin et al., 1995). This present

study's correlation result for depression and discrimination ability was similar to Kemperman et al's (1997) findings on a sample of borderline personality disorder patients. However, further detailed analysis was not possible because Kemperman et al. (1997) did not report the correlation coefficient for the non-significant result. An inspection of the BDI scores for this present study showed that the range of scores was relatively low compared to previous studies. When compared to a normative data set of chronic pain sufferers seen at a pain management research centre, the BDI scores for the chronic pain and healthy controls group fell within the 35th and 10th percentile respectively (Nicholas et al., 2008). The percentile data is available online at http://www.pmri.med.usyd.edu.au/resources/5fc583d6523ce6d03b800dcdce942383.pdf (accessed 02 Aug 2007). This implied that the low back pain sample in mine study may be more comparable to a 'general' CLBP population than patients at a pain management research centre in terms of the level of depressive symptoms.

8.8.6 Duration of low back pain and discrimination ability

The results showed that increasing duration of low back pain was correlated to better discrimination ability ($\tau = 0.194$, $p = 0.039$). Previous studies that investigated noxious discrimination ability in CLBP did not report the association of duration of low back pain and discrimination ability. An interpretation of this finding is difficult because the research design of this present study is fairly descriptive and cross-sectional in nature. It should also be considered that although relationships between variables may be uncovered, this does not establish causality. However, the relationship between duration of low back pain and discrimination ability may warrant further investigation.

8.8.7 Analgesic medication and discrimination ability

Due to ethical considerations, participants were not requested to stop their medication before and during the conduct of the study. Therefore, in order to account for the potential influence of the analgesic medication on the discrimination ability to noxious thermal stimulation, the participants were requested to provide their 24 hour medication history for analysis. The issue of quantifying medication consumption by participants in pain studies is challenging. The judgment of equivalence of different medications is a difficult one to make and involves the type of medication, amount and dose consumed. The use of the Medication Quantification Scale captures this

information and makes an approximate judgment on the equivalence of different pain medication consumed by chronic pain sufferers in this study. In contrast, previous studies have adopted a simpler approach by keeping a log of the class of medication consumed without consideration of the dosage (Yang et al., 1985; Kemperman et al., 1997). This study found that the amount of potentially analgesic medication consumed (within 24 hours before the study) did not significantly correlate with discriminability. This meant that there is no relationship between the amount of analgesic drugs consumed 24 hours before the study and the discrimination ability of the participants. This result agrees with those found by Yang et al (1985) where no significant correlation was found between pain medication taken over the previous two weeks and noxious thermal discrimination ability.

8.9 Conclusion

This present study has found that CLBP sufferers demonstrated increased discrimination ability to noxious thermal stimuli compared to healthy individuals. This finding is in contrast to findings of previous studies which generally found lowered discrimination ability to noxious thermal stimuli. However, no significant correlation was found between the psychological variables investigated (depression and anxiety), medication consumed with noxious discrimination ability. The assertion regarding the construct validity of discriminability indicating pain perception processes in association with the variables of depression, anxiety and medication consumption for CLBP sufferers cannot be made. However, the assertion that there may be a difference in discrimination ability between CLBP sufferers and healthy individuals can be made. One possible mechanism for this observed difference may be explained by Eccleston & Crombez's (1999) cognitive-affective model of the interruptive function of pain. It is recommended here that future research investigate the possible link between attentional mechanisms and the increased noxious discrimination ability found in CLBP sufferers. This chapter addressed specific objective 4 (p.7) by examining the construct validity of discriminability as a correlate of pain perception processes associated with psychological factors (depression and anxiety) for CLBP sufferers.

Chapter 9
General discussion

9.1 Introduction

A literature review was conducted looking at previous research using signal detection theory (SDT) as the model of investigation in pain perception (Chapter 3). This was followed by a discussion of some issues posed by critics on the use of SDT as a model for studying pain perception (Chapter 4). The issue of construct validity of SDT measures as an indicator of pain perception processes was discussed. Criticisms regarding the construct validity of SDT measures were categorised into three domains: theoretical, methodological and definitional/interpretational. It was decided at this juncture that response bias would not be analysed for this thesis. The reason for this decision was based on the problem of response bias artefacts created through the use of correction methods for response categories with zero proportions.

As the studies within this thesis used the confidence-rating scale, it was important to know if the results between the confidence and magnitude-rating scales were comparable. The results of the study in Chapter 5 showed that the confidence-rating scale was comparable to the magnitude-rating scale using the analytical framework proposed by Irwin & Whitehead (1991). Through Irwin & Whitehead's (1991) framework, two additional measures of cumulative discriminability function and Weber fraction were introduced and used. These were useful for examining participant discriminability for the entire range of stimuli used within the studies compared to the conventional single data point description.

This thesis proceeded by examining the construct validity of discriminability as an indicator of analgesia induced by a topical local anaesthetic, the eutectic mixture of local anaesthetics (EMLA®) (Chapter 7). This study found that discriminability did not significantly decrease in the local anaesthetic condition as compared to the control condition. This result did not establish the construct validity of discriminability as an indicator of analgesia under a topical local anaesthetic. This thesis also examined the construct validity of discriminability as a measure associated with the psychological factors of depression and anxiety in chronic low back pain (CLBP) suffers (Chapter 8).

Although it was found that the discriminability of CLBP sufferers was significantly higher compared to healthy individuals, there were no statistically significant correlations between discriminability and either depression or anxiety. The strong assertion of the construct validity of discriminability as a measure associated with depression and anxiety was therefore not established in this study. However, the lesser assertion that there was a difference in discrimination ability between CLBP sufferers and healthy individuals could be made. Eccleston & Crombez's (1999) cognitive-affective model of the interruptive function of pain could explain the results of this study.

This chapter will discuss this thesis' findings in relation to the issues that have been outlined in previous chapters.

9.2 Interpretational ambiguity of SDT measures

One of the problems described in Chapter 4 was the interpretational ambiguity of the SDT measures when using the magnitude-rating scale. Some researchers have attempted to address this question by proposing theories for an overarching analytical framework. This framework could be used as a common analytical foundation for comparing and expounding the similarities and differences of outcomes obtained by different psychophysical procedures. The overarching framework comprised of models proposed by Durlach & Braida (1969), Braida & Durlach (1972) and Laming (1984, 1997). Irwin & Whitehead (1991) then integrated these models into the overarching framework which is utilised by this thesis for analysis in Chapters 6, 7 and 8.

The framework may be used to analyse responses from both discrimination and magnitude responses assigned to physical stimuli. It is argued that magnitude responses may be considered to be only ordinal on the hierarchy of the levels of measurement (Laming, 1997), as opposed to other investigators who interpret magnitude responses as interval or ratio in nature (Stevens, 1971; Marks, 1974). If magnitude responses are considered ordinal, analysis of the responses would be restricted to the response frequencies assigned to the categories describing the perceived magnitude of the stimuli. Therefore, using these category frequencies from

magnitude responses, the discriminability between the stimuli may be estimated. The procedure for obtaining discriminability from magnitude responses would be similar to those of discrimination methods. That is, d' is generated by using cumulated proportions (Chapter 2, Section 2.7.2). The discriminability from magnitude responses may thus be compared with the responses generated through the discrimination method. Previous studies have provided evidence that this approach is feasible for auditory intensity perception (Durlach & Braida, 1969; Braida & Durlach, 1972; Pyn, Braida & Durlach, 1972, Macmillan, Braida & Goldberg, 1987). This work has also been extended to pain perception using electrocutaneous stimuli (Irwin et al., 1994). The study within Chapter 6 has provided further evidence that this framework may be used to compare the discrimination and direct scaling methods, which corresponded to the confidence-rating and magnitude-rating scales respectively (Tan, Palmer, Martin & Roche, 2007).

One of the measures generated from Irwin & Whitehead's (1991) framework is the cumulative discriminability function. This measure was used for evaluating the participants' overall discrimination within the stimulus range for this thesis. It was suggested in this thesis, and by Irwin et al. (1994), that cumulating the adjacent d's for a range of stimuli may be useful in determining the overall performance of the participants on the task. This tool has been used to predict the influences of stimulus range, stimulus probability, and type of psychophysical task on the overall discrimination ability in pain and auditory research (Irwin & Whitehead, 1991; Irwin et al., 1994; Pyn, Braida & Durlach, 1972). Irwin et al. (1994) indeed found it useful as an observational tool for describing discrimination ability recovery from a topical local anaesthetic. In this present thesis, it was also found to be useful for the comparison of the discrimination ability of participants when two types of psychophysical tasks were presented (Chapter 6), and the discrimination ability between CLBP sufferers and healthy individuals (Chapter 8). Referring back to Figure 8.5, the figure shows the cumulative discriminability functions for the CLBP sufferers and the healthy individuals. It is clear that the overall discrimination ability of the CLBP sufferers was better than the healthy individuals. This is demonstrated by the steeper slope of the cumulative discriminability function for the low back pain group. There is a more profound theoretical implication through the use of the cumulative discriminability function. The functions implied that the d' obtained for the adjacent

temperature pairs may be summated. Therefore this meant that discrimination could be viewed as perceptual distances. That is, just as physical distances (in metres) between several spatial points may be added to obtain the total physical distance, similarly this analogy may be extended to perceptual distances. It has been defined earlier that the cumulative discriminability function is a representation of the summated perceptual ability for adjacent physical stimulus quantities (for example, between 45°C, 46°C, 47°C and 48°C). The discriminability between any two chosen physical stimulus points within the 45°C to 48°C may be obtained.

However, unanticipated judgment behaviour may be an issue with the use of cumulative discriminability function. For example, in Chapter 7, the discrimination ability in the partial local anaesthetised condition demonstrated only decreased estimated d' values for the lowest noxious temperature pair (45-46°C). This thesis suggested that, despite the observed non-statistical difference in discrimination ability for the higher temperature pairs between the local anaesthesia and control groups, partial local anaesthesia was present. The apparently undiminished discrimination ability in the anaesthetised state may be explained by the hypothesis that participants adopted an attentional switch strategy. Participants switched from discriminating noxious sensation in the control condition to residual thermal sensation in the partial local anaesthesia condition. And it was probably coincidental that the d' for residual thermal sensation (for the local anaesthetic condition) was similar to that of noxious thermal sensation (for the control condition) for the higher temperature pairs within the experimental stimuli range. Therefore, it is not the usefulness of the cumulative discriminability function that is questioned. Rather, it is the interpretation of the d' changes or the apparent lack of change that determines the salience of the findings based on cumulative discriminability function. It is recommended that cumulative discriminability function is not used in isolation for the interpretation of results. Interpretation should be contingent on several converging descriptive and inferential outcomes, for example d' and pain threshold together with cumulative d'.

Despite the goal of comparability between methods being achieved, an unanticipated issue arose. For the study reported in Chapter 6, and in a related paper based on this study (Tan et al., 2007), it was suggested that an alternative interpretation could be offered for the findings of the study. The corresponding outcomes from the

magnitude-rating and confidence-rating scales could be due to dimensional overlap. The dimensionality referred to here was not related to the perceptions induced by the physical modality (i.e. heat and thermal pain). Rather, the dimensions were related to the category descriptors of the scales used in the study. It was proposed that the correspondence of the results by both the confidence and magnitude-rating scales is superficial. Therefore, the implication is that the two scales may not be related in any theoretically profound way, contrary to propositions made by Braida & Durlach (1972) and Irwin & Whitehead (1991). The reason that dimensionality was not initially considered in this thesis was because the assumption of perceptual one-dimensionality was adopted (Braida & Durlach, 1972; Durlach & Braida, 1969; Macmillan & Creelman, 2005, pp.114-115). Perceptual one-dimensionality meant that the physical stimulus (in this case, thermal stimulus) produced an internal representation that may be described in one dimension. The original study set out in Chapter 6 was not designed to test the perceptual one-dimensionality assumption.

It was proposed that this dimensionality issue could be tested in several ways (Tan et al., 2007). One method was to analyse participant responses from both magnitude-rating and confidence-rating scales using a multidimensional approach. For example, Clark, Carroll, Yang & Janal (1986) and Clark, Ferrer-Brechner, Janal, Carroll & Yang (1989) have analyzed the dimensions of both experimental and clinical pain using Individual Differences Scaling (INDSCAL) procedures. Another method would be to observe the directional shifts of discriminability from both confidence and magnitude-rating scales when an analgesic or anaesthetic procedure has been performed (Rollman, 1983). If the anaesthetic procedure leads to similar directional shifts in discriminability for both tasks, this provides some evidence that responses from both tasks exist on similar dimensions. A disconfirmation test for the perceptual dimension similarity issue may also be investigated on painful clinical conditions. That is, some characteristics of the painful condition may interact with the experimental stimulus, which then yields opposite shifts in discriminability between confidence-rating and magnitude-rating methods. Even so, disconfirmation does not negate the potential usefulness of both tasks for diagnostic purposes. In fact the underlying basis for the opposite shifts in discriminability, be it biological or cognitive in nature, could be elucidated and applied as a powerful clinical diagnostic tool for painful conditions.

9.3 Operationalisation of *d′* using research context

This thesis has shown, in some ways, that it is possible to manipulate the research design of the study to operationalise the construct of *d′* to represent either sensory or non-sensory factors. This is advantageous because of the potentially broad application of SDT for the study of the different dimensions and facets of pain. This thesis showed that this broad application is possible with appropriate operationalisation of constructs and research designs. There was inconclusive evidence from this thesis to suggest that *d′* is a sole indicator of sensory factors (Chapter 7) or psychological factors (Chapter 8) because of results that contrasted with previous research and predicted outcomes. Nevertheless, it was demonstrated in Chapter 8 that it may be possible to describe CLBP sufferers' perceptual performance using SDT methodology. However, it is unknown whether the higher *d′* found for chronic low back pain sufferers compared to healthy individuals was due to affective variables. This is because the correlations of *d′* with the affective variables examined were non-significant. For the study involving the use of a topical local anaesthetic, it cannot be established whether *d′* for the study represented the sensory function of the participants (Chapter 7). However, based on anecdotal reports from participants that residual thermal sensation was still present after the local anaesthetic intervention, it is conjectured that participants could have discriminated the post anaesthetic stimuli based on residual thermal sensations. If future studies provide evidence that the residual thermal sensation is used for judgements by participants, the *d′* obtained would likely indicate the sensory ability of the participants to discriminate between innocuous thermal sensations. Therefore, the research context and manipulation of research design is the key factor in operationalising the construct of *d′*.

9.4 The correspondence between *d′* and nociception

Perhaps the more difficult question to answer is whether discrimination ability provided information about nociception or does discrimination ability simply illuminate properties of the discrimination task (Wolff, 1986). The latter position that *d′* provides information only about discrimination is adopted by Rollman (1977). This question is very closely linked to an implicit assumption held within this present thesis. It was assumed that since the discrimination task was performed within the

noxious range of the thermal stimulus, the judgments made by participants would be based on the noxious intensity rather than the thermal intensity. This assumption is perhaps also shared implicitly by previous investigators since their interpretation of *d'* is related to changes in the perception of pain (Clark, 1994; Goolkasian, 1983; Irwin & Whitehead, 1991; Nahra & Plaghki, 2005). There is some evidence from attentional studies to suggest that participants will be more likely to focus on the noxious stimuli rather than the innocuous thermal intensity of the stimuli. Crombez et al. (1996, 1998a, 1998b) have shown that noxious stimuli were more attentionally demanding compared to other innocuous stimuli when both were presented together. This could have led to the participants performing poorer on the outcome of the innocuous task in Crombez and colleagues' studies. The innocuous tasks usually involved the participants performing an attentional task, for example the Stroop test, memorising items on a list and tonal discrimination (Crombez, 2006; Crombez, Baeyens & Eelen, 1994; Crombez et al., 1996, 1997, 1998a, 1998b; Crombez, Vervaet, Baeyens, Lysens & Eelen, 1996). The explanation forwarded by Eccleston & Crombez (1999) to explain this phenomenon was that pain is attention demanding and interruptive of other concurrent attentional processes. This meant that once the participant's attention is engaged onto the noxiousness of the stimuli, it is fairly difficult to disengage it from the noxious stimuli (Eccleston & Crombez, 1999). For this thesis, this phenomenon implied that it was very likely that participants made discriminative judgments based on the noxious rather than the innocuous thermal intensities.

In Chapter 7, the explanation of participants performing an attentional switch from noxious sensations to innocuous thermal sensations was provided to explain the discriminability results when partial skin anaesthesia was present. Assuming that this conjecture was descriptive of the cognitive processes involved in the study, this may be indicative of further evidence for the attentional interruptive-demand hypothesis by Eccleston & Crombez (1999). The perceived noxiousness of the stimuli was diminished to become innocuous by the topical local anaesthetic. This probably led to the loss of the attention demanding characteristics of the stimuli. Since the noxious stimuli were no longer attentionally engaging, the participant was able to switch their attention towards discriminating the thermal sensation. However, this account of a potential cognitive mechanism for noxious discrimination processes requires further study before any definite conclusions can be made.

Based on the above findings, it may be said that discrimination ability did provide some information about nociception, however, it required that the data be interpreted in the context under which nociception was examined. For most situations where the experimental stimuli or clinical pain experienced is deemed to be suprathreshold in nature, then it would be reasonable to assume that the d' obtained indicated the discrimination of noxious intensity. However, it is still inconclusive as to whether d' indicated a reduction in discriminative ability for noxious intensities when the intervention caused a change in stimuli strength from perceived noxiousness to innocuousness.

9.5 Affect-cognition involvement in discrimination ability for chronic pain

Numerous factors have been investigated for their relationship to chronic pain development. These include both affective as well as cognitive factors (Fernandez, 2002; Flor & Turk, 2006). For this thesis, two specific affective factors were examined in Chapter 8: depression and anxiety. The choice of these two factors was based on previous SDT pain studies where most studies concentrated on depression and anxiety. Interestingly, neither factors correlated significantly with noxious thermal discrimination ability in CLBP sufferers for this study. In contrast, the results of this thesis did not support the findings in these previous studies. There is some evidence to suggest that depression may be associated with certain cognitive deficits or biases that will ultimately be reflected in poorer discrimination ability (Brébion, Smith & Widlocher, 1997; Dworkin et al., 1995). Magnetic resonance scans have also shown altered brain activation in chronic pain sufferers. These alterations in brain region activation may be associated with cognitive biases demonstrated through functional cognitive tasks (Apkarian et al., 2004a; Apkarian et al., 2004b; Villemure & Bushnell, 2002). In contrast, chronic low back pain sufferers, for the study in Chapter 8, were found to display higher noxious thermal discrimination ability when compared to healthy individuals. This thesis proposes that this could be due to potential selection bias inherent in the recruitment of the chronic low back pain sufferers for that study. The CLBP sample was drawn from a musculoskeletal outpatients physiotherapy whereas other studies have drawn their patient sample from

psychiatric clinics or specialised pain rehabilitation units. It is recommended that future SDT studies examine the correlation between discriminability and affective factors for patient within a specialised pain rehabilitation unit. These results for patients recruited from specialised pain rehabilitation units could be compared with those recruited from general musculoskeletal outpatient clinics. Another strategy is to adopt a quasi-experimental design in which the independent variables are the affective-cognitive factors. This would examine how these factors and their magnitude may be associated with discrimination ability of noxious stimuli. Some examples of cognitive factors that may be worth exploring within the context of discrimination ability are self-efficacy (Bandura, 1977; Keefe, Rumble, Scipio, Giordano, & Perri, 2004), catastrophising (Sullivan et al., 2001) and vigilance-related concepts (Crombez, 2006).

9.6 Response bias: The unexamined variable

Signal detection theory yields two indices: a discriminability and response bias index. This thesis has focused on the discriminability index (d'). The exclusion of the response bias index within the results of this study was due to the creation of response bias artefacts through the use of correction methods for response categories with zero proportions.

In the generation of the SDT measuress for rating procedures, it was often found that participants did not use all of the available six rating categories available on the response set. Since response bias is calculated based on the z scores of the hit and false alarm rates, zero values for either of these will produce z scores of infinity. This will, in turn, produce nonsensical results. One method to overcome this extreme proportion problem is the elimination of categories that contain the offending extreme value by merging it with an adjacent category as outlined in Chapter 3. Although this correction method solved the problem of nonsensical results, it made comparison of response bias between participants impossible. The response set of 6 categories would produce 5 response bias outcomes if all the categories were utilised. However, if one category was collapsed to correct for extreme proportion, then only 4 response biases would be available for analysis. The number of response biases generated for all participants within one study could vary between 2-5 outcomes, depending on which

categories contained the extreme proportions. This made comparison undesirable because the response biases may not be generated from the same adjacent stimulus pairs. Therefore, it was decided that response bias would not be analysed in order to avoid creating analytical artefacts.

9.7 Effect of feedback on SDT measures

Clark (2007, personal communication) raised an issue regarding the effect of feedback on the participants' response bias shifts. The issue was that if feedback was provided, the participant would not hold the response biases stable for judging the stimuli. Instead, they would shift their response bias in order to optimise their chances of obtaining an accurate response in the next trial. This variability in response bias, due to inconsistency of the participant in the use of a decision cut-off, may cause a decrease in d'. This potential decrease in d' is caused mainly through the addition of an unknown amount of variance into the participants' internal representation of the stimuli (Wickelgren, 1968). Clark's (2007, personal communication) main point was not about response bias itself. Rather, it was the potential for feedback to influence response bias variability, and in turn, indirectly impact upon the discriminability. However, little is known about the contamination of participants' variable response biases on estimated d' values (Macmillan & Creelman, 2005, p.46).

This raises the question over the desirability of feedback, and the circumstances of feedback inclusion or exclusion in the procedures. This would probably be dependent on the aims of an experiment. Macmillan & Creelman (2005, pp.129-130) stated that the answer to this question is contingent on the physical stimulus intensity range and presentation probabilities of the stimuli in an experiment. When participants respond to stimuli, it is possible that judgments made may be compared with a weighted average of the stimulus effects. That is, the overall outcome of the participants' performances is due to comparison of the perceived stimuli to an averaged reference of all the stimuli. This is known as the adaptation level effect (Helson, 1964). Therefore, if the range of the stimuli is wide, then the adaptation level effect may be more prominent. Or, similarly, if the presentation probability of the stimuli is higher for a particular stimulus, then the participants' averaged reference would shift towards the more frequent stimulus. This effect has been further extended in Parducci's (1974)

range-frequency model. If the adaptation level effects are expected from participants, then the use of feedback would decrease the chances of the effect happening. In the case of this thesis, the temperature range is restricted to the two temperatures that were being tested for that test block. As for presentation probability, the stimuli were equally probable in their presentation. Therefore, it was unlikely that range-frequency effects were prominent in influencing d' values. This meant that the inclusion or exclusion of feedback within the study procedures did not adversely influence the results.

For the studies in this thesis, trial-by-trial feedback was provided for a different reason. Feedback was provided because participants tended to use a restricted range of the response categories. In order to reassure the participants and to encourage the use of the entire range of categories, feedback was given. If variability in response bias, through the use of feedback, did decrease d', then the estimated d' values in this thesis would have been lower compared to other similar studies. However, a comparison found that the d' values did not deteriorate through the use of feedback when compared to other similar studies without feedback (see Chapter 8, Figure 8.8). This comparison took account of the differences in temperature for the stimuli pairs used in those studies. This provided some evidence that feedback did not greatly decrease the d' values for the results of this thesis.

9.8 Limitations

This section will outline several limitations, both theoretical and methodological, identified within this thesis.

9.8.1 Perceptual one-dimensionality assumption

Models are used in research for conceptualising, describing, explaining and predicting physical or psychological phenomena (Graziano & Raulin, 2007, pp39-40). The analytical framework used in this thesis is a model that described and analysed participant responses obtained from the administration of noxious stimuli. The framework could be used to analyse rating experiments generating either discrimination or magnitude responses. However, inherent within the model is the assumption of perceptual one-dimensionality. As explained in an earlier section

(Chapter 6, Section 6.4), this meant that the internal representation of the stimuli may be represented as perceptual distances for the associated stimulus range. Thus these perceptual distances may be added to obtain the perceptual performance of the participant over the stimulus range. This thesis has made the implicit assumption that participants receive and integrate the stimuli inputs before generating the output as judgments. This description of dimensionality of physical stimuli should not be confused with the notion of multidimensionality of pain perception within the study of pain (Melzack & Casey, 1968; Melzack & Katz, 2006).

If the assumption of one-dimensionality is violated by participants through anomalous responding, the results generated from the analytical framework may be relatively less robust. This is because the additivity of perceptual distances may not adequately describe the estimation of d' over the stimulus range in consideration. One method of overcoming this limitation is to adopt a multidimensional model in the design and analysis of the data (MacMillan & Creelman, 2005). This multidimensional approach of designing and analysing data has been used in recognition memory and speech discrimination research (Banks, 2000; Kingston & Macmillan, 1995).

9.8.2 The variability of d'

It was observed that the variability of d' was moderately wide. This is a general issue for research involving human participants providing verbal responses related to pain perception (Clark, 1994; Hatem, Attal, Willer & Bouhassira, 2006). There are several strategies to overcome this issue. One is to increase the number of trials administered to the participants. However, this may introduce fatigue effects that may influence the participants' performance on the task. It may also increase the likelihood of hyperalgesia being induced and burns occurring. These reasons were the impetus behind the reduction of trial numbers for the studies in Chapters 6 and 7. Another reason for the choice of lower trial numbers was to reduce the total test duration per participant. A second strategy is to increase the number of trials but separate the sensory testing into several sessions. This recommendation is reasonable for purely descriptive studies or studies involving cross-sectional observation. However, this strategy may prove more challenging in implementation when applied to studies involving interventions. This is because studies involving interventions have to take

into account the effective duration of the intervention. The testing session needs to be located within the effective duration in order to capture the intervention effect.

This issue of *d′* variability is partly related to statistical power, therefore an alternative way of partially addressing this problem is to manipulate other study design parameters (for example, number of participants, effect size and participant familiarity of the task) (Lenth, 2001).

9.8.3 Averaging tendency of multiple trials

The studies within this thesis have utilised trial numbers of 17 and 40 per stimulus intensity. For each participant, this equates to between 102 (Chapter 8) and 408 (Chapter 7) trials administered for each study. This duration of trial length administration is relatively more extensive when compared to conventional studies that take between 3 to10 readings of an outcome measure for each participant. The much lengthier method used in this thesis tended to capture the mean representation of the participant's performance during the extended testing period. This is one of the reasons the reported *d′* measures were described as estimate measures. From another point of view, *d′* measures were described as estimates because the measures were calculated using hit rates and false alarm rates. Both the hit and false alarm rates are estimates of probabilities. This means that the hit and false alarm rates may vary from one block of trials to the next. Therefore, the resultant outcome of *d′* is also an estimate.

For pain perception studies that utilise relatively short-acting analgesics (for example the use of EMLA in Chapter 7) capturing the effect of the local anaesthetic within a fairly short period before the effect diminishes is important. This is because the half-life of the pharmacological product may be only a few hours in length. This is also the length of most SDT studies using a moderate number of trials in the testing procedure. This means that if the intervention is short-acting, the procedures used in SDT trials may not be appropriate for the study design. Otherwise, multiple administrations or stronger doses of the pharmacological product are required to maintain the required effect in order for testing to be carried out.

9.8.4 Examination of both pain thresholds and *d′*

In this thesis, the decision was taken not to collect data concerning the pain thresholds of the participants. The main reason for taking this decision was that procedures for the rating tasks were fairly lengthy and inclusion of the pain threshold measurements would further lengthen this testing duration. However, in retrospect, the information on pain thresholds would have been useful for three reasons. The first reason is the comparison of directionality between pain threshold and discrimination ability. Both measures could either move correspondingly in the same direction or diverge in opposite directions, depending on the context of the study. Of course, it is also entirely possible that one measure might remain static whilst the other measure changes. The implications of obtaining data for both pain threshold and discrimination ability directionality could perhaps partially answer the issue of response set dimensionality overlap outlined in Section 9.2. The second reason would be the objective verification of residual thermal responses felt by participants after a topical local anaesthetic has been administered (Chapter 7). In this situation, it would be useful to include the thermal detection threshold in addition to the thermal pain threshold. The third reason is the comparison of pain threshold change with raised discrimination ability for CLBP sufferers. In Chapter 8, the discrimination ability of CLBP sufferers were found be better than healthy individuals. The pain threshold could provide further information as to whether this finding was due to a potential selection bias for the sample group in this study. It is recommended that future studies obtain threshold measures with the intention of providing more evidence for evaluating changes in pain perception.

9.9 Recommendations for future studies

There are several recommendations based on findings from this thesis as well as thoughts about future progression of the current line of SDT pain perception research.

9.9.1 Analysis of group nociceptive discrimination behaviour

There may be research situations in which very low stimulus trial numbers may be required. Participant age or medical condition, the transient nature of the phenomenon investigated or participant intolerance of the procedure may require the application of lower trial numbers. Under such circumstances, it may still be possible to analyse the

data generated by using a research design consisting of a large number of participants and low trial numbers. One way of analysing the results is by taking the mean of all the *d′* generated for each participant. However, doing so with very low trial numbers would unnecessarily introduce large statistical bias into the outcomes (Hautus, 1997). There are two techniques in mathematical psychology that may be used to reduce the statistical bias: 1) Group Operating Characteristic Analysis (Drga, 1999) and, 2) Function of Replications Combined Estimation (FORCE) (Lapsley Miller, 1999). These techniques reduce the statistical bias by averaging out the between-subject inconsistency in the decisions made for the stimuli. Both Group Operating Characteristic Analysis and FORCE were developed within the area of auditory discrimination research. There is potential to translate these techniques into the study of pain perception. The large amount of variance encountered for participant-generated responses using experimental pain stimuli may be reduced using these techniques. Nevertheless, it could be argued that such techniques are displays of mathematical gymnastics and do not relate functionally to the variable in question. There may be some justification for this position and it is always prudent to adopt good research design rather than perform sophisticated analytical procedures to reduce unexplained variance within the data. On the other hand, if there is a firm mathematical foundation for the technique, judicious application to research should be encouraged. Some examples of mathematical techniques used in pain perception research are the algorithms used in obtaining functional magnetic resonance imaging and the analysis of evoked potentials (Logothetis et al., 2001; Chen et al., 2007; Logothetis, Pauls, Augath, Trinath & Oeltermann, 2007).

9.9.2 Attentional switching in discrimination tasks

It was noted in Chapter 7 that attentional switching could be a potential cognitive strategy used by participants in the study involving topical local anaesthetic (EMLA). Future research could explore the cognitive implications of such an attention switch. A programme of work based on this topic could establish evidence for such a cognitive mechanism. Further work could examine if there were any differences between chronic pain sufferers and healthy individuals in the performance of the attention switch. This could elucidate the cognitive processes involved in chronic pain sufferers. This programme of work would contribute to the current literature of cognitive bias in pain perception of chronic pain sufferers (Pincus & Morley, 2001).

9.9.3 Psychophysical characteristics of hyperalgesia

Another recommendation that is linked to the proposed programme on cognitive bias in chronic pain perception is the elucidation of characteristics of hyperalgesia using discrimination methods. Currently, primary and secondary hyperalgesia in healthy individuals and patient populations is examined using threshold measures (LaMotte, Thalhammer, Torebjork, & Robinson, 1982; Meyer & Campbell, 1981, O'Neill, Manniche, Graven-Nielsen & Arendt-Nielsen, 2007). Discrimination tasks may be used to investigate descriptively the performance of participants when hyperalgesia has been induced. It may be hypothesised, based on Eccleston & Crombez's (1999) cognitive-affective model of the interruptive function of pain, that noxious discrimination ability would be better for the hyperalgesic state because of the participants' heightened vigilance, cognitive engagement and perhaps neurophysiological interaction with the noxious stimuli. In contrast, the prediction could also be made that discrimination ability of the participants in the hyperalgesic state would be diminished because the sensitised neural events in either the peripheral or central sensitisation process contribute neural or perceptual 'noise'. These sources of noise could cloud the participants' discrimination clarity, thus resulting in diminished *d′*. Some early work in this area has been pursued and it was shown that secondary hyperalgesic states yield better discrimination ability (Eriksen, 2006). However it is still unclear as to the basis for such a phenomenon.

9.9.4 Noxious discrimination ability in chronic pain

In Chapter 8, it was found that CLBP sufferers had better discrimination ability compared to healthy individuals. This finding was contrary to the results of previous studies that have found CLBP sufferers having poorer noxious discrimination ability (Yang et al., 1985; Cohen et al., 1983). A novel interpretation of Eccleston & Crombez's (1999) cognitive-affective model of the interruptive function of pain was offered in Chapter 8. It might be that the chronic engagement of the chronic pain sufferer's attention on the noxious stimuli allowed the participants to better discriminate between the noxious stimuli. Further work is required to replicate this result and test the prediction. However, it is entirely possible that other mechanisms or perceptual processes may contribute to the higher noxious discrimination ability in chronic pain sufferers.

The findings of Chapter 5 provided some evidence that when the participants judged the intensities of thermal stimuli, the decisions were made within the context of the type of task and stimulus range. Since most experimental measures of pain in clinical studies have used variants of the direct scaling method, the findings in Chapter 5 suggest that the responses of clinical participants may contain a component of judgment variance. In order to prevent diminishing the participants' discrimination ability by the effect of stimulus range comparisons (Poulton, 1989), the responses of clinical participants may be examined using the discrimination method, within the framework proposed by Irwin & Whitehead (1991).

9.10 Summary

This chapter reiterated the findings of this thesis in relation to the issues outlined in previous chapters. All the issues are linked to the exploration of the construct validity of d' as an indicator of pain perception processes. The interpretational ambiguity issue was partly resolved through the utilisation of the confidence-rating response set. The comparability of the results obtained from the confidence and magnitude-rating scales are bridged by Irwin & Whitehead's (1991) analytical framework. However, this approach posed a new problem regarding the potential overlap of response set dimensions. This meant that the results obtained from the two scales may only be superficially linked and lacked a more profound theoretical basis for comparison.

Regarding the issue of the construct validity of d' in pain perception, this thesis planned two studies to investigate the idea that d' may be interpreted based on the context of the research design. Chapter 7 examined d' in the sensory context within a topical local anaesthetic state and Chapter 8 examined d' in the affective context for chronic low back pain sufferers. Based on the study in Chapter 7, the result was inconclusive in establishing that d' represented sensory function. However, the mechanism of attention switching was proposed to provide a potential explanation for the results. The results of Chapter 8 did not support the strong assertion that d', in the context of the study, represented psychological factors. However a lesser assertion could be made about the significantly higher noxious discrimination ability of the

CLBP sufferers compared to healthy individuals. This finding requires further investigation.

The following limitations of the thesis were outlined: the potential consequences of violating the perceptual one-dimensionality assumption, the averaging tendency of using multiple trials and the non-examination of both d' and pain thresholds as companion measures. Based on these limitations and other issues raised through the use of discrimination ability in this thesis, recommendations were made for future research.

9.11 Conclusion

This thesis has examined the usefulness of noxious discrimination ability as a construct for pain perception research. This was partly achieved by addressing the interpretational ambiguity of the SDT indices through the use of Irwin & Whiteheads' (1994) analytical framework. The use of the confidence-rating task reduced the number of possible interpretations for a single outcome. This provided increased clarity for interpretation of findings generated by SDT methodologies. The findings from this thesis also support the use of this framework in terms of the transferability of findings between the confidence-rating task and the magnitude-rating task. This would allow future pain research using SDT to be more easily compared with past research.

Evidence was then gathered to attempt to support the construct validity of discriminability as an indication of nociception and pain perception. The two research contexts were decreased sensory and nociceptive responses through the use of a topical local anaesthetic (EMLA®) and, verification of differences in pain perception between CLBP sufferers and healthy individuals.

The results from this thesis did not establish the usefulness of noxious discrimination ability for investigating pharmacologically-induced antinociception. The discriminability index did not reflect sensory changes induced by the topical local anaesthetic. This is in contrast to previous studies by Irwin et al. (1994) that have found significant decreases in discriminability once local anaesthesia has been induced. It is also possible that this thesis' study did not possess sufficient statistical power for detecting sensory changes brought on by the local anaesthetic.

The results for the study on CLBP sufferers found that patients had better noxious thermal discrimination compared to healthy individuals. This finding contrasted with previous research findings that patients tended to display lower discriminability. However, the discrimination ability did not correlate with either of the investigated affective constructs (depression and anxiety). This study did not establish the construct validity of discriminability as an indicator of psychological factors

(depression and anxiety) in CLBP sufferers. Further research is needed to examine and verify the findings.

This thesis has addressed some of the issues raised by critics, through theoretical, methodological and interpretational modifications, on the use of SDT measures in pain perception. This enables a more robust analysis of the construct validity issue of discriminability in pain perception. Based on the findings of Chapters 7 and 8, it is recommended that further research should continue to focus on verifying the construct validity of discriminability as an indicator of pain perception processes. This thesis recommends that caution be taken on the use of discriminability as a measure of nociceptive or pain perception changes until the construct validity issue has been satisfactorily resolved.

References

American Psychiatric Association (1987). *DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders*. (3rd ed.) Washington D.C.: American Psychiatric Press.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders IV*. (4th ed.) Washington D.C.: American Psychiatric Association.

American Psychological Association (2001). *Publication Manual of the American Pscyhological Association*. (5th ed.) Washington, D.C.: American Psychological Association.

Apkarian, A. V., Stea, R. A., & Bolanowski, S. J. (1994). Heat-induced pain diminishes vibrotactile perception: a touch gate. *Somatosensory & Motor Research, 11,* 259-267.

Apkarian, A. V., Sosa, Y., Krauss, B. R., Thomas, P. S., Fredrickson, B. E., Levy, R. E. et al. (2004a). Chronic pain patients are impaired on an emotional decision-making task. *Pain, 108,* 129-136.

Apkarian, A. V., Sosa, Y., Sonty, S., Levy, R. M., Harden, R. N., Parrish, T. B. et al. (2004b). Chronic back pain is associated with decreased prefrontal and thalamic gray matter density. *The Journal Of Neuroscience, 24,* 10410-10415.

Arendt-Nielsen, L. & Bjerring, P. (1988). Laser-induced pain for evaluation of local analgesia: a comparison of topical application (EMLA) and local injection (lidocaine). *Anesthesia And Analgesia, 67,* 115-123.

Arendt-Nielsen, L. & Bjerring, P. (1989). The effect of topically applied anaesthetics (EMLA cream) on thresholds to thermode and argon laser stimulation. *Acta Anaesthesiologica Scandinavica, 33,* 469-473.

Arntz, A., Dreessen, L., & Merckelbach, H. (1991). Attention, not anxiety, influences pain. *Behaviour Research And Therapy, 29,* 41-50.

Baird, J. C., Green, D. M., & Luce, R. D. (1980). Variability and sequential effects in cross-modality matching of area and loudness. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 277-289.

Baliki, M. N., Chialvo, D. R., Geha, P. Y., Levy, R. M., Harden, R. N., Parrish, T. B. et al. (2006). Chronic pain and the emotional brain: specific brain activity associated with spontaneous fluctuations of intensity of chronic back pain. *The Journal Of Neuroscience, 26,* 12165-12173.

Bandura, A. (1977). Self-efficay: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11,* 267-273.

Bargones, J. Y. & Werner, L. A. (1994). Adults listen selectively; infants do not. *Psychological Science, 5,* 170-174.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *Journal of Personality Assessment, 67,* 588-597.

Bekesy, G. V. (1947). A new audiometer. *Acta Oto-Laryngologica, Stockholm, 35,* 411-422.

Berglund, B. & Harju, E. L. (2003). Master scaling of perceived intensity of touch, cold and warmth. *European Journal of Pain, 7,* 323-334.

Bini, G., Cruccu, G., Hagbarth, K. E., Schady, W., & Torebjork, E. (1984). Analgesic effect of vibration and cooling on pain induced by intraneural electrical stimulation. *Pain, 18,* 239-248.

Bjerring, P. & Arendt-Nielsen, L. (1990). Depth and duration of skin analgesia to needle insertion after topical application of EMLA cream. *British Journal Of Anaesthesia, 64,* 173-177.

Bolanowski, S. J., Gescheider, G. A., Fontana, A. M., Niemiec, J. L., & Tromblay, J. L. (2001). The effects of heat-induced pain on the detectability, discriminability, and sensation magnitude of vibrotactile stimuli. *Somatosensory & Motor Research, 18,* 5-9.

Borg, G. (1998). *Borg's perceived exertion and pain scales*. Champaign, IL, US: Human Kinetics.

Braida, L. D. & Durlach, N. I. (1972). Intensity perception: II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America, 51,* 483-502.

Brébion, G., Smith, M. J., & Widlocher, D. (1997). Discrimination and response bias in memory: Effects of depression severity and psychomotor retardation. *Psychiatry Research, 70,* 95-103.

Bushnell, M. C., Duncan, G. H., Dubner, R., & He, L. F. (1984). Activity of trigeminothalamic neurons in medullary dorsal horn of awake monkeys trained in a thermal discrimination task. *Journal Of Neurophysiology, 52,* 170-187.

Bushnell, M. C., Taylor, M. B., Duncan, G. H., & Dubner, R. (1983). Discrimination of innocuous and noxious thermal stimuli applied to the face in human and monkey. *Somatosensory Research, 1,* 119-129.

Caterina, M. J., Schumacher, M. A., Tominaga, M., Rosen, T. A., Levine, J. D., & Julius, D. (1997). The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature, 389,* 816-824.

Central Register Office for Scotland (2003). *Scotland's Census 2001: Supporting information - Version 1, 13 February 2003*. Edinburgh: Central Register Office for Scotland.

Chapman, C. R. (1978). Pain: The perception of noxious events. In R.A.Sternbach (Ed.), *The psychology of pain* (pp. 169-202). New York: Raven Press.

Chapman, C. R. (1986). Pain, Perception, and illusion. In R.A.Sternbach (Ed.), *The psychology of pain* (2nd ed., pp. 153-179). New York: Raven Press.

Chapman, C. R. & Butler, S. H. (1978). Effects of doxepin on perception of laboratory-induced pain in man. *Pain, 5,* 253-262.

Chapman, C. R., Chen, A. C., & Bonica, J. J. (1977). Effects of intrasegmental electrical acupuncture on dental pain: evaluation by threshold estimation and sensory decision theory. *Pain, 3,* 213-227.

Chapman, C. R., Wilson, M. E., & Gehrig, J. D. (1976). Comparative effects of acupuncture and transcutaneous stimulation on the perception of painful dental stimuli. *Pain, 2,* 265-283.

Chapman, C. R. (1977). Sensory decision theory methods in pain research: A reply to Rollman. *Pain, 3,* 295-305.

Chapman, C. R. & Feather, B. W. (1973). Effects of diazepam on human pain tolerance and pain sensitivity. *Psychosomatic Medicine, 35,* 330-340.

Charlton, E. (1995). Ethical guidelines for pain research in humans. Committee on Ethical Issues of the International Association for the Study of Pain. *Pain, 63,* 277-278.

Chen, Z., Ohara, S., Cao, J., Vialatte, F., Lenz, F. A., & Cichocki, A. (2007). Statistical modeling and analysis of laser-evoked potentials of electrocorticogram recordings from awake humans. *Computational Intelligence and Neuroscience, 2007,* 1-25.

Clark, W. C. (1969). Sensory-decision theory analysis of the placebo effect on the criterion for pain and thermal sensitivity. *Journal Of Abnormal Psychology, 74,* 363-371.

Clark, W. C. (1974). Pain sensitivity and the report of pain: an introduction to sensory decision theory. *Anesthesiology, 40,* 272-287.

Clark, W. C. (1994). The Psyche in the Psychophysics of Pain - An Introduction to Sensory Decision Theory. *Touch, Temperature, and Pain in Health and Disease, 3,* 41-62.

Clark, W. C. (2004). Comment on: Increased pain sensitivity in fibromyalgia: effects of stimulus type and mode of presentation, Petzke et al., Pain 105 (2003) 403-413, and the related editorial: Hyperalgesia versus response bias in fibromyalgia, Fillingim, Pain, 105 (2003) 385-386. *Pain, 109,* 524-525.

Clark, W. C., Carroll, J. D., Yang, J. C., & Janal, M. N. (1986). Multidimensional scaling reveals two dimensions of thermal pain. *Journal Of Experimental Psychology: Human Perception And Performance, 12,* 103-107.

Clark, W. C., Ferrer-Brechner, T., Janal, M. N., Carroll, J. D., & Yang, J. C. (1989). The dimensions of pain: a multidimensional scaling comparison of cancer patients and healthy volunteers. *Pain, 37,* 23-32.

Clark, W. C. & Goodman, J. S. (1974). Effects of suggestion on d' and Cx for pain detection and pain tolerance. *Journal Of Abnormal Psychology, 83,* 364-372.

Clark, W. C., Janal, M. N., Zeidenberg, P., & Nahas, G. G. (1981). Effects of moderate and high doses of marihuana on thermal pain: a sensory decision theory analysis. *Journal Of Clinical Pharmacology, 21,* 299S-310S.

Clark, W. C. & Mehl, L. (1973). Signal detection theory procedures are not equivalent when thermal stimuli are judged. *Journal of Experimental Psychology, 97(2), 148-53.*

Cohen, D. J. & Lecci, L. B. (2001). Using magnitude estimation to investigate the perceptual componenets of Signal Detection Theory. *Psychonomic Bulletin & Review, 8,* 284-293.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New Jersey: Lawrence Erlbaum Associates.

Cohen, M. J., Naliboff, B. D., Schandler, S. L., & Heinrich, R. L. (1983). Signal detection and threshold measures to loud tones and radiant heat in chronic low back pain patients and cohort controls. *Pain, 16,* 245-252.

Craig, J. C. & Rollman, G. B. (1999). Somesthesis. *Annual Review of Psychology, 50,* 305-331.

Crombez, G. (2006). Hypervigilance and attention to pain: Experimental and clinical evidence. In H.Flor, E. Kalso, & J. O. Dostrovsky (Eds.), *Proceedings of the 11th World Congress on Pain* (pp. 515-528). Seattle: IASP Press.

Crombez, G., Baeyens, F., & Eelen, P. (1994). Sensory and temporal information about impending pain: the influence of predictability on pain. *Behaviour Research And Therapy, 32,* 611-622.

Crombez, G., Eccleston, C., Baeyens, F., & Eelen, P. (1996). The disruptive nature of pain: an experimental investigation. *Behaviour Research And Therapy, 34,* 911-918.

Crombez, G., Eccleston, C., Baeyens, F., & Eelen, P. (1997). Habituation and the interference of pain with task performance. *Pain, 70,* 149-154.

Crombez, G., Eccleston, C., Baeyens, F., & Eelen, P. (1998a). Attentional disruption is enhanced by the threat of pain. *Behaviour Research And Therapy, 36,* 195-204.

Crombez, G., Eccleston, C., Baeyens, F., & Eelen, P. (1998b). When somatic information threatens, catastrophic thinking enhances attentional interference. *Pain, 75,* 187-198.

Crombez, G., Vervaet, L., Baeyens, F., Lysens, R., & Eelen, P. (1996). Do pain expectancies cause pain in chronic low back patients? A clinical investigation. *Behaviour Research And Therapy, 34,* 919-925.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Davis, G. C., Buchsbaum, M. S., & Bunney, W. E., Jr. (1979). Analgesia to painful stimuli in affective illness. *The American Journal Of Psychiatry, 136,* 1148-1151.

DeCarlo, L. T. & Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General, 119,* 375-396.

Domino, G. (2002). *Psychological Testing.* New York: Cambridge University Press.

Dorfman, D. D. & Berbaum, K. S. (1986). RSCORE-J: Pooled rating-method data: A computer program for analyzing pooled ROC curves. *Behavior Research Methods, Instruments & Computers, 18,* 452-462.

Dorn, S. D., Palsson, O. S., Thiwan, S. I. M., Kanazawa, M., Clark, W. C., van Tilburg, M. A. L. et al. (2007). Increased colonic pain sensitivity in irritable bowel syndrome is the result of an increased tendency to report pain rather than increased neurosensory sensitivity. *Gut, 56,* 1202-1209.

Dougher, M. J. (1979). Sensory decision theory analysis of the effects of anxiety and experimental instructions on pain. *Journal Of Abnormal Psychology, 88,* 137-144.

Drga, V. (1999). *The theory of group operating characteristic analysis in discrimination.* Unpublished PhD thesis. University of Wellington, New Zealand.

Durlach, N. I. & Braida, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America, 46,* 372-383.

Dworkin, R. H., Clark, W. C., & Lipsitz, J. D. (1995). Pain responsivity in major depression and bipolar disorder. *Psychiatry Research, 56(2), 173-81*.

Eccleston, C. & Crombez, G. (1999). Pain demands attention: a cognitive-affective model of the interruptive function of pain. *Psychological Bulletin, 125,* 356-366.

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics, 62,* 425-451.

Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.

Egekvist, H. & Bjerring, P. (1997). Comparison of the analgesic effect of EMLA cream and EMLA patch on normal skin evaluated with laser-induced pain stimuli. *Acta Dermato-Venereologica, 77,* 214-216.

Eriksen, H. B. (2006). *How does secondary hyperalgesia affect punctate pressure discrimination in healthy subjects? A psychophysical study.* Unpublished MSc thesis. Queen Margaret University, Edinburgh.

Fernandez, E. (2002). *Anxiety, depression and anger in pain*. Dallas: Advanced Psychological Resources Inc.

Fernandez, E. & Turk, D. C. (1992). Sensory and affective components of pain: separation and synthesis. *Psychological Bulletin, 112(2), 205-17.*

Field, A. (2005). *Discovering statistics using SPSS (2nd ed.)*. Thousand Oaks, CA, US: Sage Publications, Inc.

Flor, H. & Turk, D. C. (2006). Cognitive and learning aspects. In S.B.McMahon & M. Koltzenburg (Eds.), *Wall and Melzack's Textbook of Pain* (pp. 241-258). Edinburgh: Churchill Livingstone.

Friedman, P. M., Fogelman, J. P., Nouri, K., Levine, V. J., & Ashinoff, R. (1999). Comparative study of the efficacy of four topical anesthetics. *Dermatologic Surgery, 25,* 950-954.

Fuchs, P. N., Pappagallo, M., & Meyer, R. A. (1999). Topical EMLA pre-treatment fails to decrease the pain induced by 1% topical capsaicin. *Pain, 80,* 637-642.

Gescheider, G. A. (1997). *Psychophysics: The fundamentals (3rd ed.).* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Goolkasian, P. (1983). An ROC analysis of pain reactions in dysmenorrheic and nondysmenorrheic women. *Perception & Psychophysics, 34,* 381-386.

Gracely, R. H. (2006). Studies of pain in human subjects. In S.B. McMahon & M. Koltzenburg (Eds.), *Wall and Melzack's Textbook of Pain* (3rd ed., pp. 267-304). Edinburgh: Churchill Livingstone.

Gracely, R. H., Clauw, D. J., Ambrose, K., & Petzke, F. (2004). Reply to Clark's letter. *Pain, 109,* 525-526.

Graven-Nielsen, T., Sergerdahl, M., Svensson, P., & Arendt-Nielsen, L. (2001). Methods for induction and assessment of pain in humans with clinical and pharmacological examples. In L.Kruger (Ed.), *Methods in pain research* (pp. 263-304). Boca Raton: CRC Press.

Graziano, A. M. & Raulin, M. L. (2007). *Research methods: A process of inquiry*. (6th ed.) Boston: Pearson Education Inc.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, England: Wiley.

Green, D. M., Luce, R. D., & Duncan, J. E. (1977). Variability and sequential effects in magnitude production and estimation of auditory intensity. *Perception & Psychophysics, 22,* 450-456.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424-429.

Grilly, D. M. & Genovese, R. F. (1979). Assessment of shock discrimination in rats with signal detection theory. *Perception & Psychophysics.25(6), 466-72*.

Hanley, J. A. (1988). The robustness of the "binormal" assumptions used in fitting ROC curves. *Medical Decision Making, 8,* 197-203.

Harden, R. N., Weinland, S. R., Remble, T. A., Houle, T. T., Colio, S., Steedman, S. et al. (2005). Medication Quantification Scale Version III: update in medication classes and revised detriment weights by survey of American Pain Society Physicians. *Journal of Pain. 6(6), 364-71*.

Hardy, J. D., Wolff, H. G., & Goodell, H. (1947). Studies on pain: Discrimination of differences in intensity of a pain stimulus as a basis of a scale of pain intensity. *The Journal Of Clinical Investigation, 26,* 1152-1158.

Hardy, J. D., Wolff, H. G., & Goodell, H. (1948). Studies on pain: An investigation of some quantitative aspects of the dol scale of pain intensity. *The Journal Of Clinical Investigation, 27,* 380-386.

Harkins, S. W. & Chapman, C. R. (1977). The perception of induced dental pain in young and elderly women. *Journal Of Gerontology, 32,* 428-435.

Hatem, S., Attal, N., Willer, J. C., & Bouhassira, D. (2006). Psychophysical study of the effects of topical application of menthol in healthy volunteers. *Pain, 122,* 190-196.

Hautus, M. J. & Irwin, R. J. (1992). Amplitude discrimination of sinusoids and narrow-band noise with Rayleigh properties. *Perception & Psychophysics. 52(1), 53-62*.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments & Computers, 27,* 46-51.

Hautus, M. J. (1997). Calculating estimates of sensitivity from group data: Pooled versus averaged estimators. *Behavior Research Methods, Instruments & Computers, 29,* 556-562.

Hayes, R. L., Bennett, G. J., & Mayer, D. J. (1975). Letters. *Science, 189,* 65-68.

Helson, H. (1964). *Adaptation-level theory*. Oxford, England: Harper & Row.

Hodos, W. (1970). Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74,* 351-354.

Howell, D. C. (2007). *Statistical methods for psychology*. (6th ed.) Belmont: Thomson Wadsworth.

Hoyt, D. P. & Magoon, T. M. (1954). A validation study of the Taylor Manifest Anxiety Scale. *Journal of Clinical Psychology, 10,* 357-361.

Hubley, A. M. & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *Journal of General Psychology, 123,* 207-215.

International Association of Pain Task Force on Taxonomy (1994). *Classification of pain: Descriptions of chronic pain syndromes and definitions of pain terms.* (2nd ed.) Seattle: IASP Press.

International Organisation for Standardization (2004). *International vocabulary of basic and general terms in metrology (VIM) (Draft Guide).* (3rd ed.) Geneva: International Organisation for Standardization.

Irwin, R. J., Hautus, M. J., Dawson, N. J., Welch, D., & Bayly, M. F. (1994). Discriminability of electrocutaneous stimuli after topical anesthesia: detection-theory measurement of sensitivity to painful stimuli. *Perception & Psychophysics. 55(2), 125-32.*

Irwin, R. J. & McCarthy, D. (1998). Psychophysics: Methods and analyses of signal detection. In K.A.Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 291-321). New York, NY, US: Plenum Press.

Irwin, R. J. & Whitehead, P. R. (1991). Towards an objective psychophysics of pain. *Psychological Science, 2,* 230-235.

Janal, M. N., Colt, E. W., Clark, W. C., & Glusman, M. (1984). Pain sensitivity, mood and plasma endocrine levels in man following long-distance running: effects of naloxone. *Pain, 19,* 13-25.

Janal, M. N., Glusman, M., Kuhl, J. P., & Clark, W. C. (1994). Are runners stoical? An examination of pain sensitivity in habitual runners and normally active controls. *Pain, 58(1), 109-16.*

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 92-104.

Jones, B. (1979). Signal detection theory and pain research. *Pain.7(3), 305-12.*

Kee, W. G., Steedman, S., & Middaugh, S. J. (1998). Medication quantification scale (MQS): Update of detriment weights and medication additions. *American Journal of Pain Management, 8,* 83-88.

Keefe, F. J., Rumble, M. E., Scipio, C. D., Giordano, L. A., & Perri, L. C. M. (2004). Psychological aspects of persistent pain: Current state of the science. *The Journal of Pain, 5,* 195-211.

Kemperman, I., Russ, M. J., Clark, W. C., Kakuma, T., Zanine, E., & Harrison, K. (1997). Pain assessment in self-injurious patients with borderline personality disorder using signal detection theory. *Psychiatry Research, 70(3), 175-83.*

Kingston, J. & Macmillan, N. A. (1995). Integrality of nasalization and F-sub-1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America, 97,* 1261-1285.

Laming, D. (1984). The relativity of 'absolute' judgements. *British Journal of Mathematical and Statistical Psychology, 37,* 152-183.

Laming, D. (1997). *The measurement of sensation*. New York, NY, US: Oxford University Press.

LaMotte, R. H. & Campbell, J. N. (1978). Comparison of responses of warm and nociceptive C-fiber afferents in monkey with human judgments of thermal pain. *Journal Of Neurophysiology, 41,* 509-528.

LaMotte, R. H., Thalhammer, J. G., Torebjork, H. E., & Robinson, C. J. (1982). Peripheral neural mechanisms of cutaneous hyperalgesia following mild injury by heat. *The Journal Of Neuroscience, 2,* 765-781.

Lapsley Miller, J. A. (1999). *The role of the bandwidth-duration product (WT) in the detectability of diotic signals.* Unpublished PhD thesis. University of Wellington, New Zealand.

Lautenbacher, S., Moltner, A., Lehmann, W. P., & Galfe, G. (1989). SDT analysis of experimental thermal pain, with 'signal' and 'no-signal' being determined psychophysically. *Perceptual And Motor Skills, 68,* 1019-1030.

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician, 55,* 187-193.

Lindblom, U. (2005). Sensory quantification and pain. In H.Merskey, J. D. Loeser, & R. Dubner (Eds.), *The paths of pain* (pp. 253-270). Seattle: IASP Press.

Lineberry, C. G. & Kulics, A. T. (1978). The effects of diazepam, morphine and lidocaine on nociception in rhesus monkeys: A signal detection analysis. *Journal of Pharmacology and Experimental Therapeutics, 205,* 302-310.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412,* 150-157.

Macmillan, N. A. & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.)*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Macmillan, N. A., Braida, L. D., & Goldberg, R. F. (1987). Central and peripheral processes in the perception of speech and nonspeech sounds. In M.E.H.Schouten (Ed.), *The psychophysics of speech perception* (pp. 28-45). Dordrecht, Netherlands: Martinus Nijhoff Publishing.

Macmillan, N. A. & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and 'nonparametric' indexes. *Psychological Bulletin, 107,* 401-413.

Macmillan, N. A. & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York, NY, US: Cambridge University Press.

Macmillan, N. A. & Creelman, C. D. (1996). Triangles in ROC space: History and theory of 'nonparametric' measures of sensitivity and response bias. *Psychonomic Bulletin & Review, 3,* 164-170.

Macmillan, N. A. & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98,* 185-199.

Malow, R. M. (1981). The effects of induced anxiety on pain perception: a signal detection analysis. *Pain, 11(3), 397-405*.

Malow, R. M. & Dougher, M. J. (1979). A signal detection analysis of the effects of transcutaneous stimulation on pain. *Psychosomatic Medicine, 41(2), 101-8*.

Malow, R. M., West, J. A., & Sutker, P. B. (1989). Anxiety and pain response changes across treatment: sensory decision analysis. *Pain, 38,* 35-44.

Markowitz, J. & Swets, J. A. (1967). Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. *Perception & Psychophysics, 2,* 91-97.

Marks, L. E. (1974). *Sensory processes: The new psychophysics*. New York: Academic Press.

Masters, S. S., Middaugh, S. J., Kee, W. G., Carson, D. S., Harden, R. N., & Miller, M. C. (1992). Chronic-pain medications: equivalence levels and method of quantifying usage. *Clinical Journal of Pain, 8(3), 204-14.*

Mattsson, U., Cassuto, J., Jontell, M., Jonsson, A., Sinclair, R., & Tarnow, P. (1999). Digital image analysis of erythema development after experimental thermal injury to human skin: effect of postburn topical local anesthetics (EMLA). *Anesthesia And Analgesia, 88,* 1131-1136.

McBurney, D. H. (1975). Letters. *Science, 189,* 65-68.

McBurney, D. H. (1976). Signal detection theory and pain. *Anesthesiology, 44(4), 356-9.*

McCreery, D. B. & Bloedel, J. R. (1978). A critical examination of the use of signal detection theory in evaluating a putative analgesic-transcutaneous electrical nerve stimulation. *Sensory Processes, 2,* 38-57.

McDonnell, M. & Warden-Flood, A. (2000). Effect of partial anaesthesia on normal gait. *Australian Journal of Physiotherapy, 46,* 115-120.

McNicol, D. (1972). *A primer of signal detection theory*. London: Allen & Unwin.

McNicol, D. (2005). *A primer of signal detection theory*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Melzack, R. & Casey, K. L. (1968). Sensory, motivational, and central control determinants of pain: A new conceptual model. In D.Kenshalo (Ed.), *The skin senses* (pp. 423-443). Springfield: Thomas.

Melzack, R. & Katz, J. (2006). Pain assessment in adult patients. In S.B.McMahon & M. Koltzenburg (Eds.), *Wall and Melzack's Textbook of Pain* (5th ed., pp. 291-304). Edinburgh: Churchill Livingstone.

Meyer, R. A. & Campbell, J. N. (1981). Evidence for two distinct classes of unmyelinated nociceptive afferents in monkey. *Brain Research, 224,* 149-152.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Miller, J. (1996). The sampling distribution of d'. *Perception & Psychophysics, 58,* 65-72.

Møiniche, S., Dahl, J. B., & Kehlet, H. (1993). Time course of primary and secondary hyperalgesia after heat injury to the skin. *British Journal Of Anaesthesia, 71,* 201-205.

Murdock, B. B. J. & Ogilvie, J. C. (1968). Binomial variability in short-term memory. *Psychological Bulletin, 70,* 256-260.

Nahra, H. & Plaghki, L. (2005). Innocuous skin cooling modulates perception and neurophysiological correlates of brief CO2 laser stimuli in humans. *European Journal of Pain, 9(5), 521-30*.

Naliboff, B. D., Cohen, M. J., Schandler, S. L., & Heinrich, R. L. (1981). Signal detection and threshold measures for chronic back pain patients, chronic illness patients, and cohort controls to radiant heat stimuli. *Journal Of Abnormal Psychology, 90,* 271-274.

Nicholas, M. K., Asghari, A., & Blyth, F. M. (2008). What do the numbers mean? Normative data in chronic pain measures. *Pain, 134,* 158-173.

Novy, D. M., Nelson, D. V., Goodwin, J., & Rowzee, R. D. (1993). Psychometric comparability of the State-Trait Anxiety Inventory for different ethnic subpopulations. *Psychological Assessment, 5,* 343-349.

O'Neill, S., Manniche, C., Graven-Nielsen, T., & Arendt-Nielsen, L. (2007). Generalized deep-tissue hyperalgesia in patients with chronic low-back pain. *European Journal of Pain, 11,* 415-420.

Ominsky, A. J. (1979). On the possible painful consequences of misapplying signal-detection theory. *Anesthesiology, 51(6), 491-2*.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,* 407-418.

Parducci, A. (1974). Contextual effects: A range-frequency analysis. In E.C.Carterette & M. P. Friedman (Eds.), *Handbook of Perception: Vol. 2. Psychophysical judgment and measurement* (pp. 127-141). New York: Academic Press.

Pastore, R. E., Crawley, E. J., Berens, M. S., & Skelly, M. A. (2003). 'Nonparametric' A′ and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review, 10,* 556-569.

Pedersen, J. L. & Kehlet, H. (1998a). Hyperalgesia in a human model of acute inflammatory pain: a methodological study. *Pain, 74,* 139-151.

Pedersen, J. L. & Kehlet, H. (1998b). Secondary hyperalgesia to heat stimuli after burn injury in man. *Pain, 76,* 377-384.

Pertovaara, A., Martikainen, I. K., Hagelberg, N., Mansikka, H., Nagren, K., Hietala, J. et al. (2004). Striatal dopamine D2/D3 receptor availability correlates with individual response characteristics to pain. *European Journal of Neuroscience.,* 1587-1592.

Peters, M. L., Vlaeyen, J. W. S., & Kunnen, A. M. W. (2002). Is pain-related fear a predictor of somatosensory hypervigilance in chronic low back pain patients? *Behaviour Research And Therapy, 40,* 85-103.

Pincus, T. & Morley, S. (2001). Cognitive-processing bias in chronic pain: a review and integration. *Psychological Bulletin, 127,* 599-617.

Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America, 24,* 745-749.

Pollack, I. & Hsieh, R. (1969). Sampling variability of the area under the ROC-curve and of d'e. *Psychological Bulletin, 71,* 161-173.

Pollack, I. & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science, 1,* 125-126.

Poulton, E. C. (1989). *Bias in quantifying judgments*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Price, D. D., Hu, J. W., Dubner, R., & Gracely, R. H. (1977). Peripheral suppression of first pain and central summation of second pain evoked by noxious heat pulses. *Pain, 3,* 57-68.

Pynn, C. T., Braida, L. D., & Durlach, N. I. (1972). Intensity perception: III. Resolution in small-range identification. *Journal of the Acoustical Society of America, 51,* 559-566.

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science (New York, N.Y.), 277,* 968-971.

Robinson, C. J., Torebjork, H. E., & LaMotte, R. H. (1983). Psychophysical detection and pain ratings of incremental thermal stimuli: a comparison with nociceptor responses in humans. *Brain Research, 274,* 87-106.

Rollman, G. B. (1977). Signal detection theory measurement of pain: a review and critique. *Pain, 3(3), 187-211.*

Rollman, G. B. (1979). Signal detection theory pain measures: empirical validation studies and adaptation-level effects. *Pain, 6(1), 9-21.*

Rollman, G. B. (1983). Multiple representations of experimental pain. In J.J.Bonica, U. Lindblom, & A. Iggo (Eds.), *Advances in pain research and therapy* (pp. 865-869). New York: Raven Press.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-Difference Theory of Remembering and Knowing: A Two-Dimensional Signal-Detection Model.. *Psychological Review, 111,* 588-616.

Russ, M. J., Roth, S. D., Lerman, A., Kakuma, T., Harrison, K., Shindledecker, R. D. et al. (1992). Pain perception in self-injurious patients with borderline personality disorder. *Biological Psychiatry, 32(6),* 501-11.

Russ, M. J., Shearin, E. N., Clarkin, J. F., Harrison, K., & Hull, J. W. (1993). Subtypes of self-injurious patients with borderline personality disorder. *American Journal of Psychiatry, 150(12),* 1869-71.

Schulman, A. I. & Mitchell, R. R. (1966). Operating characteristics from yes-no and forced-choice procedures. *Journal of the Acoustical Society of America, 40,* 473-477.

Schumacher, R. & Velden, M. (1984). Anxiety, pain experience, and pain report: a signal-detection study. *Perceptual And Motor Skills, 58,* 339-349.

Simpson, A. J. & Fitter, M. J. (1973). What is the best index of detectability? *Psychological Bulletin, 80,* 481-488.

Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34-50.

Soetanto, A. L., Chung, J. W., & Wong, T. K. (2004). Gender differences in pain perception: a signal detection theory approach. *Acta Anaesthesiologica Taiwanica, 42(1), 15-22*.

Somedic A.G. (2002). *SENSELab - Thermotest™: User's Manual, Version 02*. Sweden: Somedic AG.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., & Vagg, P. R. (1983). *The Stait-Trait Anxiety Inventory for adults*. Redwood City: Mind Garden.

Staud, R., Price, D. D., Robinson, M. E., Mauderli, A. P., & Vierck, C. J. (2004). Maintenance of windup of second pain requires less frequent stimulation in fibromyalgia patients compared to normal controls. *Pain, 110,* 689-696.

Steer, R. A., Clark, D. A., Beck, A. T., & Ranieri, W. F. (1999). Common and specific dimensions of self-reported anxiety and depression: The BDI-II versus the BDI-IA. *Behaviour Research And Therapy, 37,* 183-190.

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review, 78,* 426-450.

Stevens, S. S. & Stevens, G. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New Brunswick, NJ, US: Transaction Publishers.

Sullivan, M. J., Thorn, B., Haythornthwaite, J. A., Keefe, F., Martin, M., Bradley, L. A. et al. (2001). Theoretical perspectives on the relation between catastrophizing and pain. *The Clinical Journal Of Pain, 17,* 52-64.

Swets, J. A. & Pickett, R. M. (1982). *Evaluation of diagnistic systems: Methods from signal detection theory*. New York: Academic Press.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Swets, J. A., Tanner, W. P. J., & Birdsall, T. G. (1961). Decision Processes In Perception. *Psychological Review, 68,* 301-340.

Tan, C. W., Palmer, S. T., Martin, D. J., & Roche, P. A. (2007). Detection theory analysis of scaling and discrimination tasks: Responses to noxious thermal stimuli. *Perception & Psychophysics, 69,* 994-1001.

Tanner, W. P. J. & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61,* 401-409.

Tillman, D. B., Treede, R. D., Meyer, R. A., & Campbell, J. N. (1995). Response of C fibre nociceptors in the anaesthetized monkey to heat stimuli: estimates of receptor depth and threshold. *The Journal Of Physiology, 485 (Pt 3),* 753-765.

Treede, R. D., Meyer, R. A., Raja, S. N., & Campbell, J. N. (1995). Evidence for two different heat transduction mechanisms in nociceptive primary afferents innervating monkey skin. *The Journal Of Physiology, 483 (Pt 3),* 747-758.

Tritrakarn, T., Lertakyamanee, J., Koompong, P., Soontrapa, S., Somprakit, P., Tantiwong, A. et al. (2000). Both EMLA and placebo cream reduced pain during extracorporeal piezoelectric shock wave lithotripsy with the Piezolith 2300. *Anesthesiology, 92,* 1049-1054.

Van Hees, J. & Gybels, J. (1981). C nociceptor activity in human nerve during painful and non painful skin stimulation. *Journal Of Neurology, Neurosurgery, And Psychiatry, 44,* 600-607.

Villemure, C. & Bushnell, M. C. (2002). Cognitive modulation of pain: how do attention and emotion influence pain processing? *Pain, 95,* 195-199.

Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology, 5,* 102-122.

Wickens, T. D. (2002). *Elementary signal detection theory*. New York, NY, US: Oxford University Press.

Wolff, B. B. (1986). Behavioral measurement of human pain. In R.A.Sternbach (Ed.), *The psychology of pain* (2nd ed., pp. 121-151). New York: Raven Press.

Yang, J. C., Clark, W. C., & Janal, M. N. (1991). Sensory decision theory and visual analogue scale indices predict status of chronic pain patients six months later. *Journal Of Pain And Symptom Management, 6,* 58-64.

Yang, J. C., Clark, W. C., Ngai, S. H., Berkowitz, B. A., & Spector, S. (1979). Analgesic action and pharmacokinetics of morphine and diazepam in man: an evaluation by sensory decision theory. *Anesthesiology, 51,* 495-502.

Yang, J. C., Richlin, D., Brand, L., Wagner, J., & Clark, W. C. (1985). Thermal sensory decision theory indices and pain threshold in chronic pain patients and healthy volunteers. *Psychosomatic Medicine, 47,* 461-468.

Yarnitsky, D., Sprecher, E., Zaslansky, R., & Hemli, J. A. (1995). Heat pain thresholds: normative data and repeatability. *Pain, 60,* 329-332.

Appendix A
Table for converting proportions to z-scores

p′	*p*	*z(p′) = -z(p)*	*p′*	*p*	*z(p′) = -z(p)*
0.001	0.999	3.090	0.21	0.79	0.806
0.002	0.998	2.878	0.22	0.78	0.772
0.003	0.997	2.748	0.23	0.77	0.739
0.004	0.996	2.652	0.24	0.76	0.706
0.005	0.995	2.576	0.25	0.75	0.674
0.006	0.994	2.512	0.26	0.74	0.643
0.007	0.993	2.457	0.27	0.73	0.613
0.008	0.992	2.409	0.28	0.72	0.583
0.009	0.991	2.366	0.29	0.71	0.553
			0.30	0.70	0.524
0.01	0.99	2.326	0.31	0.69	0.496
0.02	0.98	2.054	0.32	0.68	0.468
0.03	0.97	1.881	0.33	0.67	0.440
0.04	0.96	1.751	0.34	0.66	0.412
0.05	0.95	1.645	0.35	0.65	0.385
0.06	0.94	1.555	0.36	0.64	0.358
0.07	0.93	1.476	0.37	0.63	0.332
0.08	0.92	1.405	0.38	0.62	0.305
0.09	0.91	1.341	0.39	0.61	0.279
0.10	0.90	1.282	0.40	0.60	0.253
0.11	0.89	1.227	0.41	0.59	0.228
0.12	0.88	1.175	0.42	0.58	0.202
0.13	0.87	1.126	0.43	0.57	0.176
0.14	0.86	1.080	0.44	0.56	0.151
0.15	0.85	1.036	0.45	0.55	0.126
0.16	0.84	0.994	0.46	0.54	0.100
0.17	0.83	0.954	0.47	0.53	0.075
0.18	0.82	0.915	0.48	0.52	0.050
0.19	0.81	0.878	0.49	0.51	0.025
0.20	0.80	0.842	0.50	0.50	0.000

Appendix B
Derivation of the relative judgmental variance equation

According to Durlach and Braida (1969) and the standard model of signal detection theory, the discriminability for discrimination method may be obtained by:

$$d'_D = \alpha / \sigma_D$$

Where d'_D is the discriminability between the two adjacent classes of stimuli in the discrimination method, α is the difference between the means of the normal probability densities for these two stimuli and, σ_D is their common standard deviation. Therefore,

$$\alpha = d'_{\mathrm{D}} \cdot \sigma_D . \qquad \text{(Equations B.1)}$$

Similarly, the discriminability obtained between two stimuli for direct scaling methods is:

$$d'_S = \frac{\alpha}{\left(\sigma_D^2 + \sigma_S^2\right)^{\frac{1}{2}}}$$

where d'_S is the discriminability between the two adjacent classes of stimuli in the direct scaling method, σ_D^2 is the stimulus variance associated with the discrimination method, and σ_S^2 is the judgmental variance associated with the direct scaling method (Macmillan & Creelman, 2005, p.134). Therefore:

$$\alpha = d'_S \cdot \left(\sigma_D^2 + \sigma_S^2\right)^{\frac{1}{2}} . \qquad \text{(Equations B.2)}$$

Combining Equations B.1 and B.2,

$$d'_D \cdot \sigma_D = d'_S \cdot \left(\sigma_D^2 + \sigma_S^2\right)^{\frac{1}{2}}$$

$$\frac{d'_{\mathrm{D}}}{d'_{\mathrm{S}}} = \frac{\left(\sigma_D^2 + \sigma_S^2\right)^{\frac{1}{2}}}{\sigma_D}$$

$$\left(\frac{d'_{\mathrm{D}}}{d'_{\mathrm{S}}}\right)^2 = \frac{\sigma_D^2 + \sigma_S^2}{\sigma_D^2}$$

$$\left(\frac{d'_{\mathrm{D}}}{d'_{\mathrm{S}}}\right)^2 = 1 + \frac{\sigma_S^2}{\sigma_D^2}$$

Therefore, an estimation of the relative variance obtained for discrimination and direct scaling methods is (Macmillan & Creelman, 2005, p.134):

$$\frac{\sigma_S^2}{\sigma_D^2} = \left(\frac{d'_{\mathrm{D}}}{d'_{\mathrm{S}}} \right)^2 - 1 .$$

Appendix C
Invitation letter to participants (Chapter 8)

Queen Margaret University College

EDINBURGH

Queen Margaret University College
Physiotherapy Subject Area
Duke Street
EH6 8HF
Edinburgh
United Kingdom
Telephone: XXXXXXXX
Email: ctan@qmuc.ac.uk

6th July 2004

Dear (Last name of participant),

Re: Invitation to Participate in Clinical Research, Physiotherapy Subject Area, Queen Margaret University College

I am a physiotherapy research student with Queen Margaret University College. My research team is currently conducting a study that will compare the pain-sensing ability of people with low back pain and healthy individuals. Your name and address was obtained from the Physiotherapy Department, Western General Hospital because you are registered with the clinic. You are being asked to participate in our study because your particulars match our initial screening criteria.

The goal of our study is to find out how emotional states affect the pain-sensing ability of people. This will help health professionals develop more powerful assessment and treatment techniques in the future. The study will involve you filling in some forms and also to undergo a heat sensation testing session. The heat sensation test has been tried and tested on participants within our institution and it is a very safe procedure. The whole process will take only about a maximum of 2 hours.

Attached to this letter are an information sheet for the study and a response slip. If you are interested in participating in this research, please fill in the response slip and send it back to me using the stamped addressed envelope included in with this letter. You can also contact me via a telephone call (tel: XXXXXXXX) or email me (email: ctan@qmuc.ac.uk) to indicate your interest. If you have further questions regarding the study, please do not hesitate to contact me via phone, mail or email.

Thank you for your attention.

Yours sincerely,
Chee-Wee Tan

Appendix D
Information sheet for participants (Chapter 8)

Queen Margaret University College

EDINBURGH

SUBJECT INFORMATION SHEET

Title of study:
Comparison of the pain-sensing ability and emotional factors between low back pain patients and healthy individuals.

Introduction

There are many factors that may influence a person's ability to sense pain. Some of the factors are the emotional state of the person, the gender of the person and any medical condition that the person might have. Knowing a person's ability to sense pain is important because it will assist the health professional in planning for the care of the person. It will also benefit the development of new treatment techniques for the person in pain. This study will be looking at how a person with low back pain responds to painful stimuli and how the emotional factors affect the way s/he responds.

Study Aims:
1. To compare the pain-sensing ability of low back pain patients and healthy individuals.
2. To draw a connection between a person's emotional state and his/her pain-sensing ability.

Study Procedure:

This study will take approximately 2 hours. You will be required to do 2 tasks during the study: (1) form filling and (2) testing of your pain-sensing ability.

You will be given 3 forms to fill in that requests some general information about you and your emotional states. After that, your pain-sensing ability will be tested.

During the testing, you will be asked to place the forearm of your writing hand ON TO a warm heat-plate. The temperature will be controlled by the tester. There will be a total of 4 temperatures tested. They are grouped in pairs of 45°C -46°C, 46°C -47°C, 47°C -48°C. The sensation of the temperatures will be similar to touching the outside

surface of a moderately hot cup of tea. The duration of the forearm placement is only 3 seconds. If you are not able to tolerate the pain from the heat stimulus, you are allowed to take away your forearm before the 3 seconds is up. Each temperature pair will be presented 34 times for comparison. There are 3 temperature pairs and a total of 102 presentations will be made. The repeated presentation is to obtain a profile of your pain-sensing ability. The whole procedure, including the form-filling, is estimated to take about 2 hours.

Confidentiality of the participant

All information obtained from you will be confidential and used anonymously for research purposes only. You may choose to withdraw from the study at any stage without any reasons given. With your permission as indicated on the consent form, your GP will be informed of your involvement in this research.

If you require further information, please do not hesitate to request it from the investigator:

Principal investigator:
Chee Wee, Tan
Queen Margaret University College
Duke St,
Leith
Edinburgh
EH6 8HF
Tel: 0131 317 3665

If you wish to speak to an independent person for further advice regarding the study, you may approach the following person:

Dr Shea Palmer
Queen Margaret University College
Duke Street
Leith
Edinburgh
EH6 8HF
Tel: 0131 317 3640

Appendix E
Participant Demographics Questionnaire (Chapter 8)

Official use only: Participant no.

Personal Details

1 What is your sex?

☐ Male ☐ Female

2 What is your date of birth?

Day Month Year

3 What is (was) the full title of your main job?

4 Which of these qualifications do you have?
Tick all the qualifications that apply

☐ 'O' Grade, Standard Grade, Intermediate 1, Intermediate 2, GCSE, CSE, Senior Certificate *or equivalent*

☐ Higher Grade, CSYS, Scottish Group Award at Higher, 'A' Level, AS Level, Advanced Senior Certificate *or equivalent*

☐ GSVQ/SVQ Level 1 or 2, SCOTVEC/National Certificate Module, BTEC First Diploma, City and Guilds Craft, RSA Diploma *or equivalent*

☐ HNC, HND, SVQ Level 4 or 5, RSA Higher Diploma *or equivalent*

☐ First Degree, Higher Degree

☐ Professional Qualifications (for example, teaching, accountancy)

☐ None of these

5 Are you currently employed?

☐ Yes ☐ No If not, how long have you been unemployed

wks mths yrs

Your General Practitioner's Details

6 Who is your General Practitioner?

7 What is the name of the GP practice?

8 What are the contact details of the GP practice?

Address **Postcode**

Telephone

1 CWT v 3.0 17092004

Brief Medical History

6 Do you have any medical conditions (excluding low back pain)?
If yes please provide details below.

	Condition/s	How long have you had this condition?	Are you on medication for this condition?	
a		*wks mths yrs*	☐ Yes ☐ No	*If you are on medication for any of these conditions, please answer Question 9.*
b		*wks mths yrs*	☐ Yes ☐ No	
c		*wks mths yrs*	☐ Yes ☐ No	
d		*wks mths yrs*	☐ Yes ☐ No	
e		*wks mths yrs*	☐ Yes ☐ No	
f		*wks mths yrs*	☐ Yes ☐ No	
g		*wks mths yrs*	☐ Yes ☐ No	
h		*wks mths yrs*	☐ Yes ☐ No	
		wks mths yrs	☐ Yes ☐ No	
		wks mths yrs	☐ Yes ☐ No	

Only for Persons with Back Pain

7 How long have you had the back pain?

wks mths yrs

8 Are you experiencing back pain <u>right now</u> at this very moment in time?

☐ Yes ☐ No

9 What medication are you currently taking?

	Name of medication	What is the prescribed dosage for this medication?	How often do you take this medication?	Have you taken this medication within the last 24 hours?
a			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
b			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No

Continued next page

	Name of medication	What is the prescribed dosage for this medication?	How often do you take this medication?	Have you taken this medication within the last 24 hours?
c			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
d			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
e			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
f			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
g			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
h			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
i			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
j			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
k			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No
l			times/day, times/week, *or* ☐ taken only when required	☐ Yes ☐ No

CWT v 3.0 17092004

Appendix F
State-Trait Anxiety Inventory – Forms Y1 and Y2

SELF-EVALUATION QUESTIONNAIRE **STAI Form Y-1**

Please provide the following information:

Name_____ Date_____ S_____

Age_____ Gender (*Circle*) **M** **F** T_____

DIRECTIONS:

A number of statements which people have used to describe themselves are given below. Read each statement and then circle the appropriate number to the right of the statement to indicate how you feel *right* now, that is, *at this moment*. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

	NOT AT ALL	SOMEWHAT	MODERATELY SO	VERY MUCH SO
1. I feel calm	1	2	3	4
2. I feel secure	1	2	3	4
3. I am tense	1	2	3	4
4. I feel strained	1	2	3	4
5. I feel at ease	1	2	3	4
6. I feel upset	1	2	3	4
7. I am presently worrying over possible misfortunes	1	2	3	4
8. I feel satisfied	1	2	3	4
9. I feel frightened	1	2	3	4
10. I feel comfortable	1	2	3	4
11. I feel self-confident	1	2	3	4
12. I feel nervous	1	2	3	4
13. I am jittery	1	2	3	4
14. I feel indecisive	1	2	3	4
15. I am relaxed	1	2	3	4
16. I feel content	1	2	3	4
17. I am worried	1	2	3	4
18. I feel confused	1	2	3	4
19. I feel steady	1	2	3	4
20. I feel pleasant	1	2	3	4

STAIP-AD Test Form Y
www.mindgarden.com

SELF-EVALUATION QUESTIONNAIRE

STAI Form Y-2

Name_____ Date_____

DIRECTIONS

A number of statements which people have used to describe themselves are given below. Read each statement and then circle the appropriate number to the right of the statement to indicate how you *generally* feel. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe how you generally feel.

	ALMOST NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
21. I feel pleasant	1	2	3	4
22. I feel nervous and restless	1	2	3	4
23. I feel satisfied with myself	1	2	3	4
24. I wish I could be as happy as others seem to be	1	2	3	4
25. I feel like a failure	1	2	3	4
26. I feel rested	1	2	3	4
27. I am "calm, cool, and collected"	1	2	3	4
28. I feel that difficulties are piling up so that I cannot overcome them	1	2	3	4
29. I worry too much over something that really doesn't matter	1	2	3	4
30. I am happy	1	2	3	4
31. I have disturbing thoughts	1	2	3	4
32. I lack self-confidence	1	2	3	4
33. I feel secure	1	2	3	4
34. I make decisions easily	1	2	3	4
35. I feel inadequate	1	2	3	4
36. I am content	1	2	3	4
37. Some unimportant thought runs through my mind and bothers me	1	2	3	4
38. I take disappointments so keenly that I can't put them out of my mind	1	2	3	4
39. I am a steady person	1	2	3	4
40. I get in a state of tension or turmoil as I think over my recent concerns and interests	1	2	3	4

STAIP-AD Test Form Y
www.mindgarden.com

Appendix G
Beck Depression Inventory II

BDI-II

Date:

Name: _____ Marital Status: _____ Age: _____ Sex: _____

Occupation: _____ Education: _____

Instructions: This questionnaire consists of 21 groups of statements. Please read each group of statements carefully, and then pick out the **one statement** in each group that best describes the way you have been feeling during the **past two weeks, including today**. Circle the number beside the statement you have picked. If several statements in the group seem to apply equally well, circle the highest number for that group. Be sure that you do not choose more than one statement for any group, including Item 16 (Changes in Sleeping Pattern) or Item 18 (Changes in Appetite).

1. Sadness
- 0 I do not feel sad.
- 1 I feel sad much of the time.
- 2 I am sad all the time.
- 3 I am so sad or unhappy that I can't stand it.

2. Pessimism
- 0 I am not discouraged about my future.
- 1 I feel more discouraged about my future than I used to be.
- 2 I do not expect things to work out for me.
- 3 I feel my future is hopeless and will only get worse.

3. Past Failure
- 0 I do not feel like a failure.
- 1 I have failed more than I should have.
- 2 As I look back, I see a lot of failures.
- 3 I feel I am a total failure as a person.

4. Loss of Pleasure
- 0 I get as much pleasure as I ever did from the things I enjoy.
- 1 I don't enjoy things as much as I used to.
- 2 I get very little pleasure from the things I used to enjoy.
- 3 I can't get any pleasure from the things I used to enjoy.

5. Guilty Feelings
- 0 I don't feel particularly guilty.
- 1 I feel guilty over many things I have done or should have done.
- 2 I feel quite guilty most of the time.
- 3 I feel guilty all of the time.

6. Punishment Feelings
- 0 I don't feel I am being punished.
- 1 I feel I may be punished.
- 2 I expect to be punished.
- 3 I feel I am being punished.

7. Self-Dislike
- 0 I feel the same about myself as ever.
- 1 I have lost confidence in myself.
- 2 I am disappointed in myself.
- 3 I dislike myself.

8. Self-Criticalness
- 0 I don't criticize or blame myself more than usual.
- 1 I am more critical of myself than I used to be.
- 2 I criticize myself for all of my faults.
- 3 I blame myself for everything bad that happens.

9. Suicidal Thoughts or Wishes
- 0 I don't have any thoughts of killing myself.
- 1 I have thoughts of killing myself, but I would not carry them out.
- 2 I would like to kill myself.
- 3 I would kill myself if I had the chance.

10. Crying
- 0 I don't cry anymore than I used to.
- 1 I cry more than I used to.
- 2 I cry over every little thing.
- 3 I feel like crying, but I can't.

_____ Subtotal Page 1

Continued on Back

THE PSYCHOLOGICAL CORPORATION®
Harcourt Brace & Company
SAN ANTONIO
Orlando • Boston • New York • Chicago • San Francisco • Atlanta • Dallas
San Diego • Philadelphia • Austin • Fort Worth • Toronto • London • Sydney

0154018392

11. Agitation

0 I am no more restless or wound up than usual.
1 I feel more restless or wound up than usual.
2 I am so restless or agitated that it's hard to stay still.
3 I am so restless or agitated that I have to keep moving or doing something.

12. Loss of Interest

0 I have not lost interest in other people or activities.
1 I am less interested in other people or things than before.
2 I have lost most of my interest in other people or things.
3 It's hard to get interested in anything.

13. Indecisiveness

0 I make decisions about as well as ever.
1 I find it more difficult to make decisions than usual.
2 I have much greater difficulty in making decisions than I used to.
3 I have trouble making any decisions.

14. Worthlessness

0 I do not feel I am worthless.
1 I don't consider myself as worthwhile and useful as I used to.
2 I feel more worthless as compared to other people.
3 I feel utterly worthless.

15. Loss of Energy

0 I have as much energy as ever.
1 I have less energy than I used to have.
2 I don't have enough energy to do very much.
3 I don't have enough energy to do anything.

16. Changes in Sleeping Pattern

0 I have not experienced any change in my sleeping pattern.
1a I sleep somewhat more than usual.
1b I sleep somewhat less than usual.
2a I sleep a lot more than usual.
2b I sleep a lot less than usual.
3a I sleep most of the day.
3b I wake up 1–2 hours early and can't get back to sleep.

17. Irritability

0 I am no more irritable than usual.
1 I am more irritable than usual.
2 I am much more irritable than usual.
3 I am irritable all the time.

18. Changes in Appetite

0 I have not experienced any change in my appetite.
1a My appetite is somewhat less than usual.
1b My appetite is somewhat greater than usual.
2a My appetite is much less than before.
2b My appetite is much greater than usual.
3a I have no appetite at all.
3b I crave food all the time.

19. Concentration Difficulty

0 I can concentrate as well as ever.
1 I can't concentrate as well as usual.
2 It's hard to keep my mind on anything for very long.
3 I find I can't concentrate on anything.

20. Tiredness or Fatigue

0 I am no more tired or fatigued than usual.
1 I get more tired or fatigued more easily than usual.
2 I am too tired or fatigued to do a lot of the things I used to do.
3 I am too tired or fatigued to do most of the things I used to do.

21. Loss of Interest in Sex

0 I have not noticed any recent change in my interest in sex.
1 I am less interested in sex than I used to be.
2 I am much less interested in sex now.
3 I have lost interest in sex completely.

_____ Subtotal Page 2
_____ Subtotal Page 1
_____ Total Score

Appendix H
Published Material

105. Comparison of context-coding noise within discrimination and intensity rating methods in the study of pain perception.

CW Tan, S Palmer, D Martin, K Kirk.
Physiotherapy Subject Area, School of Health Sciences, Queen Margaret University College, Duke Street, Edinburgh, UK, EH6 8HF.

Introduction: Signal Detection Theory (SDT) (Swets, 1996) is a psychophysical theory that uses statistical methods to obtain the sensitivity of a subject separate from his/her response bias. SDT methodology employs discrimination rating instead of the conventional intensity rating (e.g. Verbal Rating Scale), for investigating pain report. Within sensory judgement, sensations are usually compared to the previous experiences of the subject before a judgement is made about it. "Context-coding noise" is the subject's decreased ability to accurately compare the sensation to the sensory experience or the context (Durlach and Braida, 1969; Irwin et al, 1994). Durlach and Braida proposed a model that showed discrimination rating to be more precise than intensity rating because it yields less context-coding noise. Therefore, the method with more context-coding noise can be ascertained by comparing the cumulated sensitivities (i.e. adding the adjacent sensitivity outcomes) along the defined stimuli-ranges of the stimulus-modality. The implication of using a less "noisy" method is to provide the researcher with a more accurate picture of the subject's sensory judgement.

Aim of investigation: To compare discrimination rating against intensity rating and determine which method yields less context-coding noise.

Method: Approval was obtained from QMUC Ethics Committee. Following informed consent, 6 subjects were recruited via convenient sampling from the staff and students of QMUC (age range=21-35, 2 males & 4 females). Equipment: MSA Thermotest (Somedic AG). Procedure: Noxious thermal stimuli were used.The single-interval, 6-categories confidence-rating discrimination and intensity-rating scaling task were compared. For the discrimination task, each subject was administered 3 blocks of trials, each block containing one stimuli-pair (45°C and 46°C, 46°C and 47°C, 47°C and 48°C). For the intensity rating task, each subject was administered one block of trials containing the same temperatures (45°C, 46°C, 47°C and 48°C). The sequences of temperature and task presentations were randomised. Forty trials of each temperature were presented. The total number of trials were 240 and 160 for the discrimination and intensity rating task respectively. The stimuli were applied to the subjects' forearm. Each stimulus lasted 3 seconds. The temperature presented was confirmed to the subject at the end of each trial for the optimisation of the subject's accuracy in judgement.

Results: The SDT index of sensitivity, d′, was computed for both tasks. Repeated measures ANOVA revealed no significant differences between the sensitivities of the 3 stimuli-pair for the discrimination task ($p=0.484$) but a significant difference for the intensity rating task ($p=0.009$). The discrimination task had a higher cumulated sensitivity score (summation of sensitivity scores) than the intensity rating task (Paired t-test, $p=0.004$).

Discussion: Using the same theoretical analysis framework, the discrimination task was shown to contain less context-coding noise than the intensity rating task. The use of cumulated sensitivity also provided evidence that a range of sensitivity performances may provide more information regarding the subject's sensory judgement than a single sensitivity score. **Conclusion:** This study provided further empirical evidence to support the use of SDT as a psychophysical model for the study of pain.

References:

Durlach NI, Braida LD (1969). Intensity perception. I. Preliminary Theory of Intensity Resolution. J Acoust Soc Am, 46(2)(Pt 2), 372-383.

Irwin RJ et al (1994). Discriminability of electrocutaneous stimuli after topical anesthesia: Detection-theory measurement of sensitivity to painful stimuli. Perception and Psychophysics, 55(2), 152-132.

Swets JA (1996). Signal Detection Theory & ROC analysis in psychology & diagnostics: Collected papers, Lawrence Erlbaum, New Jersey.

Tan, C.W., Palmer, S., Martin, D. & Kirk, K. (2004). Comparison of context-coding noise within discrimination and intensity rating methods in the study of pain perception. The Pain Society Annual Scientific Meeting, Manchester, March 30-April 2, p105.

23. The influence of stimulus presentation frequency on noxious thermal discrimination: a methodological study.

Authors: TAN CW[1], PALMER S[1], MARTIN D[2], ROCHE P[1]

[1]School of Health Sciences, Queen Margaret University College.
[2]Centre for Health & Social Care Research, Sheffield Hallam University.

[Introduction] For noxious thermal discrimination tasks, stimuli are presented repeatedly to determine the perceptual discrimination ability of subjects. In a laboratory setting, high stimulus presentation frequencies are usually utilised. However, such an endeavour may not be possible in clinical research. It is likely that a reasonable estimate of the subjects' discrimination ability may be obtained with appropriate corrections to the response data (Hautus, 1997). This study compared two stimulus frequencies to determine if the corrections will cause undesirable bias on the discrimination ability measure. The relative efficiency of tasks was also computed to describe the statistical variability of the two stimulus frequencies. **[Method]** Queen Margaret University College Research Ethics Committee approved this study. Six healthy subjects were recruited for this study. The two stimulus presentation frequencies were chosen based on a literature review of signal detection theory (SDT) studies investigating thermal pain perception. For studies using parametric and nonparametric SDT measures, the median stimulus frequencies were 40 and 17 per stimulus intensity respectively. These were considered representative of the upper-bound and lower-bound median frequencies used for past studies. The study consisted of 2 sessions: one session presented stimulus frequency at 17 per stimulus intensity (N_{17}) and one session presented stimulus frequency at 40 per stimulus intensity (N_{40}). Within each session, the experimenter tested subjects' ability to discriminate between two thermal intensities. A total of 3 potentially noxious temperature pairs (45°C & 46°C, 46°C & 47°C, 47°C & 48°C) were tested. The thermal stimuli were applied for 3 seconds to the subjects' dominant forearm using a Peltier contact thermode. The sessions and the temperature pairs within the sessions were randomised. The subjects' confidence ratings of whether the presentation was the higher or lower of each temperature pair were obtained. **[Results]** The signal detection theory index, d_a, was used to represent discrimination ability. A 2 x 3 (stimulus frequencies x temperature pairs) repeated measures ANOVA test was used to analyse the data. There were no significant differences in d_a between the conditions ($F(2,10)=0.934$, $p=0.425$). The N_{40} had a lower d_a variance than N_{17} ($var(N_{17}) = 0.513$, $var(N_{40}) = 0.295$), and was therefore relatively more efficient (relative efficiency = 0.57). **[Discussion & Conclusion]** The lower stimulus frequency did not introduce significant bias to the discrimination ability measure. It was unexpected that the N_{40} was relatively more efficient than the N_{17}. Theoretically, variability should increase with more subject responses being sampled (Hautus & Lee, 1998). A reasonable estimate of the discrimination ability was obtained by the N_{17} with a slight loss in efficiency.

References

Hautus MJ (1997). Calculating estimates of sensitivity from group data: Pooled versus averaged estimators. *Behav Res Methods Instruments Computers, 29,* 556-562.

Hautus MJ & Lee AJ (1998). The dispersions of estimates of sensitivity obtained from four psychophysical procedures: Implications for experimental design. *Percept & Psychophysics, 60,* 638-649.

Tan, C.W., Palmer, S., Martin, D. & Roche, P. (2005). The influence of stimulus presentation frequency on noxious thermal discrimination: A methodological study. The Pain Society Annual Scientific Meeting, Edinburgh, March 8-11, p23.

22. The effects of a topical anaesthetic (EMLA®) on noxious thermal discrimination: a psychophysical study.

Authors: TAN CW[1], PALMER S[1], VETO J[1], MARTIN D[2], ROCHE P[1].
[1]School of Health Sciences, Queen Margaret University College.
[2]Centre for Health & Social Care Research, Sheffield Hallam University.

[Background] The effects of topical anaesthetics have conventionally been investigated using detection threshold measures. In nociceptive systems, perception based on comparison of physical stimuli (noxious discrimination) is a commonly used perceptual but less investigated strategy. If noxious discrimination is indeed a common strategy, it is essential to investigate the behavioural outcomes of this strategy in perceiving analgesic states. This study examined the effects of a topical anaesthetic (EMLA®) using noxious thermal discrimination as the investigative procedure. **[Method]** Queen Margaret University College Research Ethics Committee approved this study. Ten healthy subjects were recruited for this study. The study consisted of 2 consecutive sessions: pre- and post-60 minute application of EMLA® to the forearm of the subjects. Both forearms were randomised for EMLA® application or control. Within each session, the experimenter tested subjects' ability to discriminate between two thermal intensities. A total of 3 potentially noxious temperature pairs (45°C & 46°C, 46°C & 47°C, 47°C & 48°C) were tested. The thermal stimuli were applied for 3 seconds to the subjects' forearm using a Peltier contact thermode. Every temperature pair was presented 34 times. The subjects' confidence ratings of whether the presentation was the higher or lower of each temperature pair were obtained. **[Results]** The subjects' discrimination ability was estimated by the signal detection theory parameter, d_a (Macmillan & Creelman, 2005). A 2 x 2 x 3 (forearms x pre-post EMLA x temperature pairs) repeated measures ANOVA test was used to analyse the data. There was no significant effect of EMLA® on noxious thermal discrimination ability ($F(2,16) = 2.263$, $p = 0.136$). **[Discussion]** The results could be due to the thermal energy penetrating beyond the depth of anaesthesia (Arendt-Nielsen & Bjerring, 1989). Although superficially located nociceptors may have been anaesthetised (Bjerring & Arendt-Nielsen, 1990), the thermal energy may have activated warmth receptors or nociceptors located deep in the dermis to allow discrimination of the stimuli. It is proposed that future studies either match thermal exposure to depth of anaesthesia by varying the spatial and temporal parameters of the thermal stimuli, or use a more superficial source of heat stimuli (e.g. lasers). However, it is also possible that the statistical test may not have sufficient power (due to low subject numbers) to detect a change in sensitivity. **[Conclusion]** The effects of EMLA® did not significantly reduce noxious discrimination ability. The stimulus type and parameters for noxious thermal discrimination, as a technique, must be taken into consideration when designing topical anaesthetic studies.

Reference

Arendt-Nielsen L & Bjerring P (1989). The effect of topically applied anaesthetics (EMLA® cream) on thresholds to thermode and argon laser stimulation. *Acta Anaesthesiol Scand, 33,* 469-473.

Bjerring P & Arendt-Nielsen L (1990). Depth and duration of skin analgesia to needle insertion after topical application of EMLA cream. *Br J Anaesth, 64,* 173.

Macmillan NA & Creelman CD (2005). *Detection theory: A user's guide*. (2nd ed.) New Jersey: Lawrence Erlbaum Associates.

Tan, C.W., Palmer, S., Martin, D., Veto, J. & Roche, P. (2005). The effects of a topical anaesthetic (EMLA) on noxious thermal discrimination: A psychophysical study. The Pain Society Annual Scientific Meeting, Edinburgh, March 8-11, p22.

Perception & Psychophysics
2007, 69 (6), 994-1001

Detection theory analysis of scaling and discrimination tasks: Responses to noxious thermal stimuli

CHEE-WEE TAN
Queen Margaret University, Edinburgh, Scotland

SHEA T. PALMER
University of the West of England, Bristol, England

DENIS J. MARTIN
University of Teesside, Middlesbrough, England

AND

PATRICIA A. ROCHE
Queen Margaret University, Edinburgh, Scotland

This study's main purpose was to examine the sensitivity estimates obtained from scaling and discrimination approaches for nociception assessment in healthy individuals. This investigation may inform future applications in diagnostic procedures for painful conditions. Models of psychophysical judgment based on those of Durlach and Braida (1969), Laming (1984), and Irwin and Whitehead (1991) were used as the common analytical framework. Noxious thermal contact stimuli were used. The results show that the scaling approach produced lower detection theory sensitivity estimates than did the discrimination approach. The additional judgment variance in scaling tasks could explain this lowered sensitivity. The relative judgmental variance value of 2.18 obtained in this study is lower than variance values found in previous investigations. This discrepancy is probably due to the relatively smaller stimulus range employed in this study. The authors propose that the theoretical framework used in this study may be used in future studies to investigate the different dimensions of pain perception.

Psychophysical measurements are commonly used to quantify the judgments of noxious stimuli. There are two main psychophysical approaches in pain research. The first, the *direct scaling method*, involves having participants estimate their perception of the intensity of the noxious stimulus. An example of such an approach using a bounded scale is the visual analog scale (Price, 1994). The second, the *discrimination method*, involves having participants discriminate between two stimuli of different noxious intensities. An example is the yes–no experiment, in which the participant is required to state whether the stronger (or more noxious) stimulus was presented. Both direct scaling and discrimination methods are used frequently in pain research—for example, in brain imaging studies (Pertovaara et al., 2004), neurophysiological studies (Nahra & Plaghki, 2005), and clinical studies (Kemperman et al., 1997). The topic of interest for our research group is the potential use of psychophysical methods as differential diagnostic tools for painful conditions. For example, a clinical scenario may consist of two similar patient groups, within a broader patient grouping, manifesting different responses to scaling or discrimination methods. This may provide the basis for further studies investigating whether the differences involve biological, affective, or cognitive dissimilarities (Petersen & Rowbotham, 2006). In order to proceed with this program, the responses of healthy individuals, as measured by scaling and discrimination methods intended for clinical application, need to be compared. However, critics have argued that the discrimination method may not quantify the perceived noxiousness of the stimuli; rather, it may only compare the perceived intensity magnitudes of the stimuli (Craig & Rollman, 1999; Rollman, 1977). Discrimination methods may therefore lack construct validity in measuring the perceived noxiousness of stimuli.

To investigate whether discrimination methods do measure the perceived noxiousness of the stimuli and whether their measurements are comparable with those of the direct scaling methods, Irwin and Whitehead (1991) and Irwin, Hautus, Dawson, Welch, and Bayly (1994) used signal detection theory as a common analytical framework for the data obtained from both methods. The index used

C.-W. Tan, ctan@qmu.ac.uk

Tan, C-W., Palmer, S.T., Martin, D.J. & Roche, P.A. (2007). Detection-theory analysis of scaling and discrimination tasks: Responses to noxious thermal stimuli. Perception & Psychophysics, 69(6):994-1001.

for estimating the sensitivity was d'. The theories, underlying the analytical framework, were originally proposed by Braida and Durlach (1972) and Laming (1984, 1997). This analytical framework provided a means of comparing different psychophysical methods in sensory perception.

Irwin et al. (1994) found that the psychometric function obtained by the discrimination method was similar to that obtained by the direct scaling method. This provided some evidence to support the use of discrimination in the measurement of noxious stimuli perception. However, the sensitivities obtained by the direct scaling method were reduced compared with those obtained by the discrimination method. This may be partially explained by the way judgments are made for both methods. Laming (1984) argued that the magnitude that participants assigned to physical stimuli may be considered to be only nominal. Therefore, information obtainable from these judgments is limited to the response frequencies assigned to different stimuli. Using the information from the response frequencies, the discriminability between the stimuli may be estimated. From this perspective, the sensitivity estimate obtained through sensory judgment is a rough estimate of sensory discrimination. According to Braida and Durlach's (1972) theory, the reduction in d' obtained by the direct scaling method resulted from the way participants judged the stimuli presented during the task. For the direct scaling method, participants compared the sensation of a given stimulus with the range of sensations of all the stimuli presented in the course of the experiment (Braida & Durlach, 1972). In addition to that, the responses made by participants for the direct scaling method may be autocorrelated to the previous response (Laming, 1984). These potential sources of judgment variance may degrade the d' for the direct scaling method. These judgment variances may not be present, or may be minimized, in the discrimination method.

Braida and Durlach's (1972) theory states that it is possible to estimate the relative extent of the judgment variance inherent in the direct scaling method compared with the variance associated with the discrimination method. According to the standard model of signal detection theory, the sensitivity is $d'_D = (\mu_2 - \mu_1)/\sigma_D$, where d'_D is the discriminability between the two adjacent classes of stimuli, μ_1 and μ_2 are the means of the normal probability densities, and σ_D is their common standard deviation. When the standard model is extended to encompass the additional variance inherent in the direct scaling method, then $d'_S = (\mu_2 - \mu_1)/(\sigma_D^2 + \sigma_S^2)^{1/2}$, where d'_S is the discriminability between the two adjacent classes of stimuli in the direct scaling method, σ_D^2 is the stimulus variance associated with the discrimination method, and σ_S^2 is the judgmental variance associated with the direct scaling method (Macmillan & Creelman, 2005, p. 134). An estimate of the additional judgmental variance relative to the stimulus variance can be obtained by

$$\sigma_S^2 / \sigma_D^2 = (d'_D / d'_S)^2 - 1 \quad (1)$$

(Durlach & Braida, 1969; Macmillan & Creelman, 2005).

One of the aims of the present investigation was to extend the analytical framework proposed by Irwin et al. (1994) to psychophysical measurements of responses to noxious thermal stimuli. The relevance of noxious thermal stimuli in pain research has been established at several levels. Such stimuli are among the most commonly used physical stimuli for evoking experimental pain (Gracely, 2005). Neurobiologically, a thermal stimulus activates a known narrow range of primary afferent fiber nociceptors—namely C-fiber and Type I and Type II A-fiber nociceptors (Meyer & Campbell, 1981; Treede, Meyer, Raja, & Campbell, 1995). At the molecular level, a noxious thermal stimulus activates a nonselective cation channel, the transient receptor potential vanilloid-1 receptor, which is a potential therapeutic target for pharmacological management of pain (Caterina et al., 1997).

Another aim was to verify previous findings by Braida and Durlach (1972), Irwin and Whitehead (1991), and Irwin et al. (1994) that the direct scaling method produced decreased d' compared with the discrimination method. The present study also estimated the relative amounts of judgment variance in both methods.

METHOD

Participants

The participants were recruited from among the students and staff of Queen Margaret University, Edinburgh, using convenience sampling. Six healthy volunteers (4 women and 2 men) took part in the experiment. The participants' median age was 28 years (range: 21–35 years).

Ethical approval. This study was approved by Queen Margaret University's research ethics committee. All of the participants provided written, informed consent for participation in this experiment.

Inclusion and exclusion criteria. The inclusion criteria were (1) age of 18 years or more and (2) ability to provide consent for participation in the study. The exclusion criteria were (1) the presence of medical conditions that caused anesthesia to the tested limb or the consumption or application of medication that caused analgesia or anesthesia to the tested limb, and (2) any wounds or injury to the tested limb.

Apparatus and Stimuli

The thermal stimuli were applied on the ventral surface of both forearms. A Thermotest (Somedic AB, Sweden) was used to administer heat stimuli via a contact thermode (with surface measuring 25 mm × 50 mm). Heating and cooling of the contact surface were achieved through a Peltier element housed within the thermode. The stimulus sets (45°C, 46°C, 47°C, and 48°C) were preprogrammed using EXPOSURE software (Somedic AB, Sweden).

Procedure

There were two tasks: a magnitude description task (MDT), representing the direct scaling method, and an intensity resolution task (IRT), representing the discrimination method. Each task was performed on different forearms for each participant, chosen at random without replacement. All randomizations within this experiment were performed using an online randomization plan generator (www.randomization.com). Each participant completed both the MDT and the IRT within the same day. Twenty practice trials, similar to the actual trials, were presented at the beginning of every task for familiarization.

Magnitude description and intensity resolution tasks. The one-interval rating task was used for both the IRT and the MDT. Each trial began with the experimenter instructing the participant to place his or her forearm on the thermode (preset at the relevant testing temperature). Each trial's observation period lasted 3 sec. An

Erratum: p995, column 1, ln 14-16, instead of "Laming (1984) argued that the magnitude that participants assigned to physical stimuli may be considered to be only *nominal*", it should have been "...considered to be *ordinal*".

automated auditory signal indicated to the participant to remove the forearm from the thermode after the 3 sec had elapsed. If participants were not able to tolerate the full length of stimulus application, they were allowed to lift their forearms away from the thermode, although no participants did so during the study. There was an interstimulus interval (ISI) of 10 sec before the next trial started.

The stimulus set for both tasks consisted of four temperatures: 45°C, 46°C, 47°C and 48°C. For the IRT trials, each trial presented one of two temperatures. There was equal probability of presentation for either of the two temperatures. There were three stimulus pairs in total: 45°C and 46°C, 46°C and 47°C, and 47°C and 48°C. The stimulus pair presentation was randomized. The three stimulus pairs of the IRT clocked a total of 240 trials per participant (80 trials for each pair). For the MDT trials, each trial presented one of four temperatures. Again, there was equal probability of presentation for any one of the four temperatures. There were a total of 160 trials per participant clocked for the MDT (40 trials for each of the four temperatures). The order of trial presentation for both tasks was randomized.

The participants verbally indicated their judgments to the experimenter, and these were recorded. For both tasks, responses were made according to response sets with six categories (see Figures 1A and 1B). The MDT required the participants to estimate the perceived magnitude of the stimulus presented according to six descriptions of sensory quality: warm, hot, faint pain, painful, very painful, and severe pain (Figure 1A). For the IRT, the participants rated their degree of confidence about whether the stimulus presented was the higher or lower intensity of a pair of stimulus intensities (Figure 1B).

The participants were told the temperature of the administered stimulus at the end of each trial—that is, trial-by-trial feedback was provided for both tasks. Participants' judgments may be biased by the comparison of observations with a weighted average of stimulus effects. This is also known as the *adaptation level effect* (Helson, 1964). Feedback was introduced to minimize this bias. An unpublished pilot study conducted by our group determined that participants were apprehensive about making judgments when no feedback was given. Feedback thus reassured the participants and encouraged the use of the entire response set.

For the MDT, the participants received the following instructions:

> In this experiment you will be asked to judge the intensities of heat stimuli presented to you. The judgment method involves assigning categories with descriptions to match the intensities of the heat sensations you will experience [Figure 1A shown to the participant]. There are six categories of intensities. Verbally indicate to the experimenter the category number with a description that matches most closely to the sensation you experienced. After you have done this, you will be told the temperature of the heat stimulus just presented to you.

For the IRT, the participants received the following instructions:

> In this experiment you will be asked to determine which one of two heat stimuli was presented to you. One stimulus is hotter than the other. Your task is to indicate whether the presented stimulus was the higher or the lower intensity and how confident you are in making that decision. There are six categories to describe your decision [Figure 1B shown to the participant]. Verbally indicate to the experimenter the category number with a description that matches most closely to your decision. After you have done this, you will be told the temperature of the heat stimulus just presented to you.

Prevention of hyperalgesia, heat injury, and windup. Two specific procedures were implemented to prevent hyperalgesia onset and heat injury of the test sites. The first procedure involved instructing the participant to shift the position of the thermode to an adjacent forearm skin area at the beginning of a new trial. The second procedure involved the enforcement of an ISI of 10 sec. The latter procedure also minimized the effect of the perceived noxiousness in latter trials increasing as a result of temporal summation. This phenomenon of noxious temporal summation is termed *windup* (Price, Hu, Dubner, & Gracely, 1977; Staud, Price, Robinson, Mauderli, & Vierck, 2004). Windup is usually maintained when the ISI is less than 3 sec. The participant's forearm was checked by the experimenter for signs of heat injury after every 20 trials or if there was a concern that heat injury might have occurred. Signs of heat injury or hyperalgesia, shown by profound erythema with pain or hypersensitivity of the skin, were identified as criteria for withdrawal from the study. No participants suffered any form of heat injury during this study.

Analysis

The receiver operating characteristic (ROC) curves of each stimuli pair and task were also plotted for every participant. The Gaussian unequal variance model was fitted to the data using the RScorePlus software written by Lewis Harvey. RScorePlus is derived from Dorf-

A

1	2	3	4	5	6
Warm	Hot	Faint Pain	Painful	Very Painful	Severe Pain

B

1	2	3	4	5	6
I am absolutely certain the weak stimulus was presented	I am fairly certain the weak stimulus was presented	I am somewhat certain that the weak stimulus was presented	I am somewhat certain that the strong stimulus was presented	I am fairly certain the strong stimulus was presented	I am absolutely certain the strong stimulus was presented

Figure 1. (A) Magnitude description scale representing the scaling method. This scale was presented to the participants during the magnitude description task (MDT) for judgment. The participants verbally provided the number that matched the description of the magnitude of sensation felt. (B) Intensity resolution scale representing the discrimination method. This scale was presented to the participants during the intensity resolution task (IRT) for judgment. The participants verbally provided the number that matched the description of their degree of confidence about which of the two stimuli (stronger or weaker) was presented.

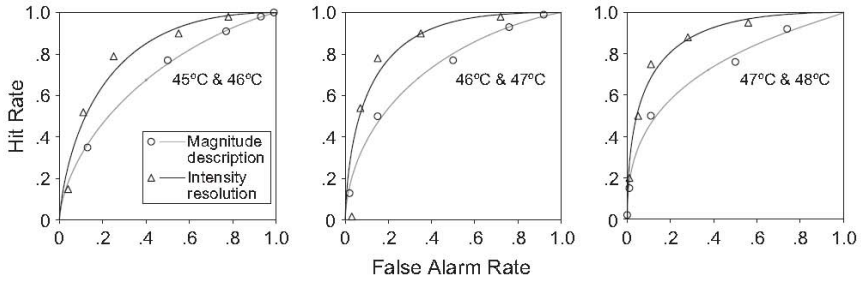

Figure 2. Receiver operating characteristic (ROC) curves fitted using a jackknifed procedure utilizing the pooled ratings of all 6 participants. Each panel shows the ROC curves of the MDT and the IRT for each stimulus pair.

man and Alf's (1969) RScore program, and it provides a maximum-likelihood fit of the signal detection model to the rating data. A total of 36 ROCs (6 participants × 3 stimuli pairs × 2 tasks) were generated for analysis. For the MDT, the adjacent temperatures were paired for analysis. This yielded the same number of stimulus pairs as did the IRT. Data from both tasks were analyzed in a similar manner. The detection theory index of discriminability, d_a (Macmillan & Creelman, 2005; Simpson & Fitter, 1973), and the slopes of the ROCs based on three stimulus pairs, s, were computed. The index d_a assumes an unequal variance model and is numerically equal to d' in the equal variance case. The Gaussian equal variance index, d', was to be adopted if s for the discrimination and scaling data did not systematically depart from unity. When extreme response frequencies were present (i.e., categories containing proportions of zero), the categories were collapsed for analysis.

Cumulative sensitivity function. The d' values of adjacent stimuli for both tasks were cumulated so that the total sensitivity across the temperature range could be visualized. Durlach and Braida (1969) named the resultant plots *cumulative sensitivity functions* (CSF). The lines of best fit through the origin were plotted using the least-squares method for the CSF of both tasks. The Weber fraction was calculated for each using the CSF. The Weber fraction, in this context, may be defined as the stimulus difference that is needed to produce a performance of $d' = 1$ as the just noticeable difference.

Relative judgmental variance. Equation 1 was used to estimate the relative variance, which is reproduced here as $\sigma^2_{MDT} / \sigma^2_{IRT} = (d'_{IRT} / d'_{MDT})^2 - 1$, where d'_{MDT} is the sensitivity between the two adjacent temperatures in the MDT, d'_{IRT} is the sensitivity between the two temperatures in the IRT, σ^2_{MDT} is the judgmental variance associated with the MDT, and σ^2_{IRT} is the stimulus variance associated with the IRT.

RESULTS

Receiver Operating Characteristics

Of the 36 total ROC curves obtained, two differed significantly from the unequal variance model at the .05 significance level according to the chi-square goodness-of-fit statistic. The individual data from all participants were jackknifed, following the approach by Dorfman and Berbaum (1986), to generate six additional ROC curves to summarize the results of all stimulus pairs in both tasks. These ROC curves are shown in Figure 2. The jackknife procedure aims to avoid the common drawbacks of conventional averaging of sensitivity estimates (Macmillan & Kaplan, 1985); one of these drawbacks is obtaining a lower estimate of sensitivity compared with the sensitivity estimates that would be obtained from the original data if no averaging was used.

The ROC slopes for the discrimination and scaling methods based on the three stimulus pairs were 1.01 (SE = .09) and 1.05 (SE = .13), respectively. The slopes for both tasks did not depart systematically from unity; therefore, the Gaussian equal variance d' was used instead of d_a.

Discriminability Results

Figure 3 summarizes the discriminability of the stimulus pairs within each task. Although the data were jackknifed to generate the ROC curves shown in Figure 2, the conventional averaging of the sensitivity means was retained in Figure 3 to show the actual data for the 6 participants. Figure 3 shows that the average discriminability of the IRT was always higher than that of the MDT. This observation is as predicted by the analytical framework. It also agrees with results from previous studies using a noxious electrocutaneous stimulus (Irwin et al., 1994; Irwin & Whitehead, 1991; Rollman, 1983). Also, the discriminability of both tasks increased with an elevation of the temperatures of the stimulus pair. A repeated measures ANOVA (2 tasks × 3 stimulus pairs) performed on the discrimination ability data showed a significant main effect of task [$F(1,5) = 24.98$, $p = .004$]. There was also a

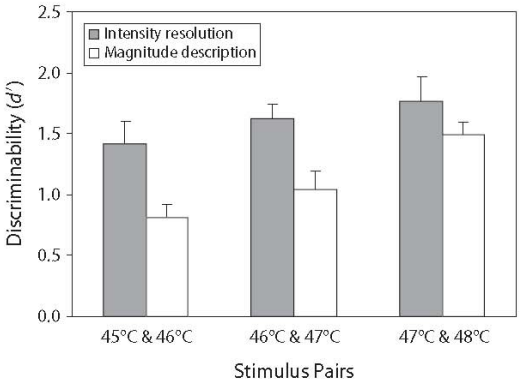

Figure 3. Mean discriminability, obtained through conventional averaging, of the MDT and IRT methods for all stimulus pairs. The error bars depict standard errors of the means.

significant main effect of stimulus pairs [$F(2,10) = 5.37$, $p = .026$]. Contrasts showed that sensitivity estimates for the 46°C–47°C stimulus pairs were not significantly higher than were those for the 45°C–46°C stimulus pairs with a large effect size [$F(1,5) = 2.63$, $p = .166$, $r = .59$]. The contrast also showed that sensitivity estimates for the 47C°–48°C stimulus pairs were significantly higher than those for the 45°C–46°C stimulus pairs with a large effect size [$F(1,5) = 7.529$, $p = .041$, $r = .60$]. However, the interaction effect between task and stimulus pair was not significant ($p = .152$).

Cumulative sensitivity functions. The CSFs were obtained using the jackknifed sensitivity estimates. The d' values of adjacent stimuli were cumulated. The successive cumulative sensitivities provided coordinates on the y-axis for plotting the CSF. Figure 4 shows the CSFs for this study. The linear functions were fitted to the data using the least-squares method, with the functions passing through the origin. There is a difference between the slopes of the two CSFs. The slope for the IRT is steeper than that of the MDT, indicating that the overall discriminability of the IRT was better than that of the MDT. Since the linear fit of these functions was adequate, it may be said that the averaged discrimination performances of the participants were in accordance with Weber's law. The Weber fractions were found to be 0.026 for the MDT and 0.015 for the IRT.

Relative judgmental variance. Using Equation 1, the additional variance in the MDT was calculated to be 2.18 times greater than the variance in the IRT. This number was calculated using the cumulated sensitivity values obtained from the CSFs for the MDT and IRT.

DISCUSSION

This study found that the MDT yielded decreased sensitivities compared with the IRT for noxious thermal stimuli. The amount of additional judgmental variance in the MDT was 2.18 times greater than that in the IRT. These results are consistent with Durlach and Braida's predictions (1969) and results from previous studies (Irwin et al., 1994; Irwin & Whitehead, 1991; Rollman, 1983).

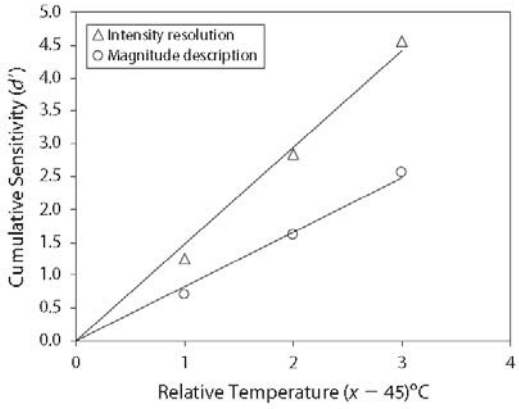

Figure 4. Cumulative sensitivity functions for the MDT and IRT as a function of relative temperature. The jackknifed d' values were used to obtain the cumulative sensitivity functions. The relative temperatures were obtained by subtracting 45°C from the higher temperature of each stimulus pair.

The Contribution of Judgmental Variance to a Poorer Sensitivity in MDT

This study was conducted to integrate the direct scaling and discrimination methods (MDT and IRT, respectively) under a common framework. The present finding of lower sensitivity estimates yielded by the MDT as compared with those yielded by the IRT supports the prediction that an additional component of variance may be attributed for direct scaling methods. Our results suggest that outcomes yielded by discrimination methods and direct scaling methods may be related. Therefore, this finding adds evidence to the assertion that discrimination methods are suitable for measuring responses from noxious stimuli. However, our results would have to be interpreted under the framework and assumptions of Durlach and Braida's (1969) theory. Irwin et al. (1994) stated that if this same analytical framework were extended to the method of magnitude estimation, similar results could be expected. It could be argued that the MDT used in this study is an example of scales that involve judgments of sensation magnitude (Braida & Durlach, 1972).

Lower Relative Variance in This Study Compared With Relative Variance in Previous Studies

It is interesting to note that the relative variance between the two methods found in the present study was 2.18, lower than that found by Irwin and Whitehead (1991). The relative variances in their description task (similar to the MDT) and identification task were 5.4 and 2.22 times more, respectively, than the variance in the discrimination task. Our result is, perhaps, not unexpected, and has two possible explanations. The first explanation may be associated with the use of relatively lower numbers of trials in this study, and the second may be connected with the stimulus range used for the MDT.

Influence of Lower Trial Numbers on Judgment Variance

The use of lower numbers of trials would inevitably increase both the variability of the responses and the likelihood of extreme proportions. This response variability may contribute considerable statistical bias to the sensitivity estimates (Hautus, 1997). An unpublished study by our research group found that when the number of trials per intensity in a one-interval confidence-rating task was decreased from 40 to 17, the amount of variance for the sensitivity estimates of the 17-trial task increased 1.74 times.

One might argue that higher numbers of trials could be used to suppress the amount of variance in the sensitivity estimates, and we acknowledge that using more trials should be done as much as is practically possible. There are, however, other factors to consider when large numbers of trials are used, such as the onset of heat injury

and hyperalgesia (Pedersen & Kehlet, 1998), the ethical acceptability of prolonged noxious stimulation (Charlton, 1995), and, ultimately, the transferability of the laboratory protocol to clinical studies. All of these factors should be carefully considered when deciding on numbers of trial presentations.

Influence of Stimulus Range on Judgment Variance

Another factor that may have influenced the amount of the judgment variance for the MDT was the range of stimuli judged. For the IRT, the participant was required to concentrate only on the difference between the two stimuli presented in a pair. This is in contrast to the MDT, which may have allowed participants to also compare the sensation magnitude of the presented stimulus to the context of the stimulus range, in spite of the trial-by-trial feedback provided to participants. A similar explanation was also offered by Rollman (1979), based on adaptation level theory (Helson, 1964). Durlach and Braida (1969) theorized that if the stimulus range for the scaling task were large, the task would become more difficult for the participants, leading to lowered sensitivity estimates. Since according to Durlach and Braida's theory, discrimination tasks are easier because the judgmental component is absent, performance on such tasks will always be better than performance on direct scaling tasks. However, Durlach and Braida predicted that for a small stimulus range, the contribution of the judgment variance in direct scaling would become almost negligible, and performance on the scaling task would be similar to that on the discrimination task. This prediction was generally supported by Pynn, Braida, and Durlach's (1972) study on auditory intensity discrimination. This raises another possible reason for the lower additional variance observed for the MDT in our investigation compared with other studies: the possibility that the relatively narrow temperature range for this study caused smaller values of the judgment variance to be found, as predicted by Durlach and Braida's theory. Nevertheless, further studies need to be conducted to confirm this conjecture.

CSF As a Potential Tool for Investigating Suprathreshold Sensitivities

The perception of noxious experimental stimuli has also been studied with methods that obtain point estimates of the transition from innocuousness to painfulness (Graven-Nielsen, Sergerdahl, Svensson, & Arendt-Nielsen, 2001). An example of the use of point estimates in pain research is the determination of the pain threshold using the method of limits. The effectiveness of pain relief treatments has been evaluated largely on the basis of the lowering of this threshold. This method does not illuminate the effects of pain relief treatments on the suprathreshold sensitivities in which pain, the construct of interest, resides. This is especially important for suprathreshold sensitivities in studies examining nociception. The same criticism could be leveled at the sensitivity estimates obtained for individual stimulus pairs in the present study. The sensitivity estimates provided information confined to only one specific stimulus pair, which reveals little about the sensitivities contained within the sensory range of interest. This problem was solved, for the purposes of this study, through the use of CSFs. CSFs may provide additional information on the suprathreshold range of sensitivities and the effects of intervention on them (Gracely, 2005), and they may be a valuable tool for future studies investigating the description and influence of interventions on suprathreshold sensitivities.

The Relevance of Our Findings for Future Clinical Studies

Our findings show that when humans judged the intensities of thermal stimuli, decisions were made within the context of the type of task and the stimulus range. Since most experimental measures of pain in clinical studies use variants of the direct scaling method, our findings suggest that the responses of clinical participants may contain a component of judgment variance. In order to prevent diminishing the participants' discrimination ability by the effect of stimuli range comparisons (Poulton, 1989), the responses of clinical participants may be examined using the discrimination method, within the framework proposed by Irwin and Whitehead (1991).

Nevertheless, clinical pain is a multidimensional experience involving affective, cognitive, and sensory components (Melzack, 1999). Chronic pain experienced by patients may be associated with changes in their empirical pain thresholds or self-reported pain intensity (Kosek, Ekholm, & Hansson, 1996). Affective, cognitive, and sensory responses to pain may interact to alter the amount of variance within the psychophysical responses. Studies have provided some evidence that signal detection theory measures of pain may be influenced by affective disorders (Dworkin, Clark, & Lipsitz, 1995; Kemperman et al., 1997). These studies have used the direct scaling method within the framework of signal detection theory. Since the direct scaling method yields an additional variance on participant responses, it would be interesting to establish, in future studies, the interaction between affect and pain response in the absence of additional variance (i.e., using the discrimination method). A clinical study examining this question in chronic pain sufferers is currently being conducted by our research group.

An Alternative Interpretation: The Dimensional Hypothesis

Although we have interpreted our findings on the basis of a theory of judgment, it is possible that the results could be due to dimensional overlap between the responses of both tasks. This would mean that the judgment theory might have to be revised for nociception, since it assumes perceptual one-dimensionality (Durlach & Braida, 1969; Macmillan & Creelman, 2005, pp. 113–115). This alternative interpretation of dimensional overlap could be tested in several ways. The first method would be to utilize a multidimensional analytical approach. For example, Clark, Yang, Carroll, and Janal (1986) and Clark, Ferrer-

Brechner, Janal, Carroll, and Yang (1989) analyzed the dimensions of both experimental and clinical pain using individual differences scaling procedures. Another method would be to observe the directional shifts of sensitivity from both discrimination and direct scaling methods when an analgesic or anesthetic procedure has been performed (Rollman, 1983). If the anesthetic procedure led to similar directional shifts in sensitivity for both tasks, this would provide some evidence that responses from both tasks existed in similar dimensions. A disconfirmation test for the perceptual dimension similarity hypothesis may also be investigated in painful clinical conditions. That is, some characteristics of the painful condition may interact with the experimental stimulus, which would then yield opposite shifts in sensitivity between discrimination and direct scaling methods. Even so, disconfirmation does not negate the potential usefulness of both tasks for diagnostic purposes. In fact, the underlying basis for the opposite shifts in sensitivity, be it biological or cognitive in nature, could be elucidated and applied as a powerful clinical diagnostic tool for future treatment of painful conditions.

CONCLUSION

This study demonstrated that the discrimination approach is comparable to the direct scaling approach. Bridging the two approaches was made possible by analyzing the data under the theoretical framework of Durlach and Braida (1969), on the basis of the assumption of perceptual one-dimensionality. Our results are consistent with Durlach and Braida's prediction that an additional component of judgment variance contributes to the decreased sensitivity in the direct scaling approach. This finding is useful for clinical pain studies that employ psychophysical methods of testing, and it may also inform diagnostic procedures for painful conditions. Regardless of the type of psychophysical method used in clinical studies, it is possible to relate and compare findings. This would also suggest that discrimination methods are admissible as psychophysical procedures for pain studies. Therefore, this framework may serve as a potentially useful tool for evaluating the often complex processes of pain perception.

AUTHOR NOTE

We thank Michael Hautus, Gary Rollman, and all the reviewers for their invaluable comments on the manuscript. Correspondence concerning this article should be addressed to C.-W. Tan, Queen Margaret University, Duke Street, Edinburgh EH6 8HF, Scotland (e-mail: ctan@qmu.ac.uk).

REFERENCES

Braida, L. D., & Durlach, N. I. (1972). Intensity perception: II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, **51**, 483-502.

Caterina, M. J., Schumacher, M. A., Tominaga, M., Rosen, T. A., Levine, J. D., & Julius, D. (1997). The capsaicin receptor: A heat-activated ion channel in the pain pathway. *Nature*, **389**, 816-824.

Charlton, E. (1995). Ethical guidelines for pain research in humans. *Pain*, **63**, 277-278.

Clark, W. C., Ferrer-Brechner, T., Janal, M. N., Carroll, J. D., & Yang, J. C. (1989). The dimensions of pain: A multidimensional scaling comparison of cancer patients and healthy volunteers. *Pain*, **37**, 23-32.

Clark, W. C., Yang, J. C., Carroll, J. D., & Janal, M. N. (1986). Multidimensional scaling reveals two dimensions of thermal pain. *Journal of Experimental Psychology*, **12**, 103-107.

Craig, J. C., & Rollman, G. B. (1999). Somesthesis. *Annual Review of Psychology*, **50**, 305-331.

Dorfman, D. D., & Alf, E., Jr. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals: Rating-method data. *Journal of Mathematical Psychology*, **6**, 487-496.

Dorfman, D. D., & Berbaum, K. (1986). RSCORE-J: Pooled rating-method data: A computer program for analyzing pooled ROC curves. *Behavior Research Methods, Instruments, & Computers*, **18**, 452-462.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **42**, 372-383.

Dworkin, R. H., Clark, W. C., & Lipsitz, J. D. (1995). Pain responsivity in major depression and bipolar disorder. *Psychiatric Research*, **56**, 181.

Gracely, R. H. (2005). Studies of pain in human subjects. In S. B. McMahon & M. Koltzenburg (Eds.), *Wall and Melzack's textbook of pain* (5th ed., pp. 267-289). Philadelphia: Churchill Livingstone.

Graven-Nielsen, T., Sergerdahl, M., Svensson, P., & Arendt-Nielsen, L. (2001). Methods for induction and assessment of pain in humans with clinical and pharmacological examples. In L. Kruger (Ed.), *Methods in pain research* (pp. 263-290). Boca Raton, FL: CRC Press.

Hautus, M. J. (1997). Calculating estimates of sensitivity from group data: Pooled versus averaged estimators. *Behavior Research Methods, Instruments, & Computers*, **29**, 556-562.

Helson, H. (1964). *Adaptation-level theory*. New York: Harper & Row.

Irwin, R. J., Hautus, M. J., Dawson, N. J., Welch, D., & Bayly, M. F. (1994). Discriminability of electrocutaneous stimuli after topical anesthesia: Detection-theory measurement of sensitivity to painful stimuli. *Perception & Psychophysics*, **55**, 125-132.

Irwin, R. J., & Whitehead, P. R. (1991). Towards an objective psychophysics of pain. *Psychological Science*, **2**, 230-235.

Kemperman, I., Russ, M. J., Clark, W. C., Kakuma, T., Zanine, E., & Harkison, K. (1997). Pain assessment in self-injurious patients with borderline personality disorder using signal detection theory. *Psychiatric Research*, **70**, 175-183.

Kosek, E., Ekholm, J., & Hansson, P. (1996). Sensory dysfunction in fibromyalgia patients with implications for pathogenic mechanisms. *Pain*, **68**, 375-383.

Laming, D. (1984). The relativity of "absolute" judgements. *British Journal of Mathematical & Statistical Psychology*, **37**, 152-183.

Laming, D. (1997). The psychophysical primitive. In *The measurement of sensation* (pp. 143-167). Oxford: Oxford University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, **98**, 185-199.

Melzack, R. (1999). From the gate to the neuromatrix. *Pain*, **82** (Suppl. 6), S121-S147.

Meyer, R. A., & Campbell, J. N. (1981). Evidence for two distinct classes of unmyelinated nociceptive afferents in monkey. *Brain Research*, **224**, 149-152.

Nahra, H., & Plaghki, L. (2005). Innocuous skin cooling modulates perception and neurophysiological correlates of brief CO_2 laser stimuli in humans. *European Journal of Pain*, **9**, 521-530.

Pedersen, J. L., & Kehlet, H. (1998). Secondary hyperalgesia to heat stimuli after burn injury in man. *Pain*, **76**, 377-384.

Pertovaara, A., Martikainen, I. K., Hagelberg, N., Mansikka, H., Nagren, K., Hietala, J., & Sheinen, H. (2004). Striatal dopamine D2/D3 receptor availability with individual response characteristics to pain. *European Journal of Neuroscience*, **20**, 1587-1592.

Petersen, K. L., & Rowbotham, M. C. (2006). Quantitative testing scaled up for multicenter clinical research networks: A promising start. *Pain*, **123**, 219-220.

Poulton, E. C. (1989). *Bias in quantifying judgments*. Hove, U.K.: Erlbaum.

Price, D. D. (1994). Psychophysical measurement of normal and abnormal pain processing. In J. Boivie, P. Hansson, & U. Lindblom (Eds.), *Touch, temperature, and pain in health and disease: Mechanisms and assessments* (pp. 3-25). Seattle: IASP Press.

PRICE, D. D., HU, J. W., DUBNER, R., & GRACELY, R. H. (1977). Peripheral suppression of first pain and central summation of second pain evoked by noxious heat pulses. *Pain*, **3**, 57-68.

PYNN, C. T., BRAIDA, L. D., & DURLACH, N. I. (1972). Intensity perception: III. Resolution in small-range identification. *Journal of the Acoustical Society of America*, **51**, 559-566.

ROLLMAN, G. B. (1977). Signal detection theory methods in pain research. *Pain*, **3**, 187-211.

ROLLMAN, G. B. (1979). Signal detection theory pain measures: Empirical validation studies and adaptation-level effects. *Pain*, **6**, 9-21.

ROLLMAN, G. B. (1983). Multiple representations of experimental pain. In J. J. Bonica, U. Lindblom, & A. Iggo (Eds.), *Advances in pain research and therapy* (pp. 865-869). New York: Raven.

SIMPSON, A. J., & FITTER, M. J. (1973). What is the best index of detectability? *Psychological Bulletin*, **80**, 481-488.

STAUD, R., PRICE, D. D., ROBINSON, M. E., MAUDERLI, A. P., & VIERCK, C. J. (2004). Maintenance of windup of second pain requires less frequent stimulation in fibromyalgia patients compared to normal controls. *Pain*, **110**, 689-696.

TREEDE, R. D., MEYER, R. A., RAJA, S. N., & CAMPBELL, J. N. (1995). Evidence for two different heat transduction mechanisms in nociceptive primary afferents innervating monkey skin. *Journal of Physiology*, **483**, 747-758.

(Manuscript received September 20, 2006;
revision accepted for publication February 5, 2007.)

graded maximal exercise test and progressively increased. Outcome measurements include BMI, functional measures (Stroke Impact Scale (SIS), Fugl-meyer, gait velocity), cardiorespiratory fitness (VO2peak, MET level, and duration of exercise test), and muscle strength (measured with a hand-held isometric dynamometer). **ANALYSIS:** An independent t-test was used to examine differences in baseline measurements between the SD and SA groups. A paired t-test was used to compare measurements before and after the exercise program in the 6 subjects who completed the intervention. Significance level was set at 0.05. **RESULTS:** At baseline, a significant difference between the groups was found for percent recovery on the SIS (37.5% for SD, 65.83% for SA, $p = 0.021$). The SD subjects also had higher BMI, decreased function (SIS score, Fugl-Meyer), decreased VO2peak, and decreased strength of all extremities compared to the SA subjects, although none of these differences were significant. For the 6 subjects who completed the intervention (3 SD, 3 SA), significant improvements were found in VO2peak (11.8 vs 14.5 ml/kg/m, $p = 0.046$), MET level (3.4 vs 4.2, $p = 0.41$), and duration of test (9.1 vs 12.9 minutes, $p = 0.037$). **CONCLUSIONS:** Subjects with both stroke and diabetes rated their level of recovery from stroke lower than subjects with stroke alone, and demonstrated generally lower levels of function. Following the intervention, significant improvements in cardiorespiratory fitness were noted even in this small sample. We have identified feasibility issues related to recruitment and a 20% drop out rate that will need to be addressed in future studies. **IMPLICATIONS:** This preliminary work indicates a need for future study with a greater number of subjects to determine if there is a difference in outcome between the two groups. The mechanism behind these hypothesized differences will be explored in the future. **KEYWORDS:** diabetes, stroke, exercise. **FUNDING ACKNOWLEDGEMENTS:** This work was supported by a Mary Switzer Fellowship award from the National Institute of Disability and Rehabilitation Research through the US Department of Education. **CONTACT:** pkluding@kumc.edu
ETHICS COMMITTEE: Human Subjects Committee of the University of Kansas Medical Center

Research Report Platform Presentation
1811 **Monday 4 June 11:05**
VCEC Meeting Room 18

NOXIOUS HEAT DISCRIMINATION ABILITY IN PERSONS WITH CHRONIC PAIN: A PSYCHOPHYSICAL STUDY

Tan C[1], Roche P[1], Martin D[2], Palmer S[3]; [1]Queen Margaret University College, School of Health Sciences, Edinburgh, United Kingdom; [2]Sheffield Hallam University, Centre for Health & Social Care, Sheffield, United Kingdom; [3]University of the West of England, Faculty of Health & Social Care, United Kingdom

PURPOSE: To compare the noxious thermal discriminative ability and decision-making bias of persons with chronic pain to pain-free individuals. **RELEVANCE:** Studies have indicated that persons with chronic pain (CP) demonstrated higher experimental pain thresholds (pressure and thermal) compared to pain-free individuals (PF). However it has not been clearly established the extent to which physiological and cognitive-decisional influences contribute to the elevated pain threshold. This information is important for understanding the effect of chronic pain on behavioural responses and the underlying neurophysiological processes of chronic pain. In particular, secondary hyperalgesia which is a consequence of central sensitisation. **PARTICIPANTS:** Thirty-three persons with chronic low back pain and twenty-nine pain-free individuals took part. They were recruited from an outpatient physiotherapy clinic and community centres. The mean age for CP was 54.7 (S.D. = 13.9) years and PF was 57.8 (S.D. = 16.7) years. The average length of low back pain experienced was 15.1 (S.D. = 10.3) years. **METHODS:** Noxious thermal stimuli were administered to the participant's dominant forearm using a Peltier thermode with a surface area of 25 mm × 50 mm (Somedic AB). Three blocks of paired temperatures were presented (45ºC and 46ºC, 46ºC and 47ºC, 47ºC and 48ºC). Participants were required to determine the higher of the paired temperature using a 6 category confidence rating scale. **ANALYSIS:** The sensitivity (da, perceptual accuracy at the task) and response bias (ca, tendency to choose a particular rating) of the participants were calculated (Macmillan and Creelman, 2005). One way repeated ANOVA with Group (CP vs. PF) as a between-subject factor was performed on the sensitivity and response bias data. **RESULTS:** CP were better at discriminating between the paired noxious temperatures than the PF as noted by their higher sensitivity scores ($p = .027$). There were no significant differences between the participant's ability to discriminate for the higher temperatures pairs as opposed to the lower temperature pairs ($p = .311$). No significant interaction effects between Group and the temperature pair ($p = .642$). CP held the same decision criteria compared to PF when asked to choose the scale ratings ($p = .24$). **CONCLUSIONS:** Interestingly, the CP group was better at discriminating noxious thermal stimuli than PF. The response biases of both groups were similar. This finding suggests that the perceptual accuracy of CP, compared to PF, in performing thermal discriminatory tasks may be enhanced. There is a current lack of models explaining this phenomenon and the relationship between pain threshold and noxious discrimination ability. However, central sensitisation of the nervous system is a potential explanatory model. **IMPLICATIONS:** Elucidating the processes underlying chronic pain may guide healthcare professionals in focusing their treatment and rehabilitation strategies (physical, psychological or cognitive). **KEYWORDS:** Chronic, pain, psychophysics. **FUNDING ACKNOWLEDGEMENTS:** This study was funded by the Queen Margaret University College PhD studentship programme. **CONTACT:** ctan@qmuc.ac.uk
ETHICS COMMITTEE: Queen Margaret University College Ethics Committee; Lothian Research Ethics Committee (National Health Service, Edinburgh)

Research Report Platform Presentation
1474 **Monday 4 June 11:05**
VCEC Meeting Rooms 19-20

EXPERIENCES OF CAREGIVERS TO PATIENTS WITH AMYOTROPHIC LATERAL SCLEROSIS REGARDING NON-INVASIVE POSITIVE PRESSURE HOME VENTILATION

Sundling I[1], Ekman S[2], Weinberg J[3], Klefbeck B[4,5]; [1]Stockholms Sodra Geriatriska Klinik, Sjukgymnastikenheten, Stockholm, Sweden; [2]Division of Gerontological Caring Science, Department of Neurobiology, Health Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden; [3]Department of Clinical Neuroscience, Karolinska University Hospital, Huddinge, Sweden; [4]Division of Physiotherapy, Department of Neurobiology, Health Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden; [5]Department of Physical Therapy, Karolinska University Hospital, Huddinge, Sweden

PURPOSE: The purpose of this study was to investigate the experiences of caregivers to patients with Amyotrophic Lateral Sclerosis (ALS), regarding non-invasive positive-pressure ventilation (NPPV) at home. **RELEVANCE:** Although strong evidence underlines the importance of treating sleep disordered breathing to improve quality of life for patients with ALS, most patients underutilize treatment. Research on quality of life during end-of-life care is sparse and interviews with caregivers may be the only source of information when patients cannot be interviewed. Little is known about the impact on the caregiver's life when a family member is provided with non-invasive positive-pressure home ventilation and no study using a qualitative approach on the experiences of caregivers to patients with ALS on home ventilation was found. **PARTICIPANTS:** Eight caregivers to patients with ALS were selected for their ability to provide information for the study. The caregivers were 40-74 years of age. All were spouses to patients with ALS, living at home. The patients had been on non-invasive ventilation at home for a

Tan, C.W., Martin, D., Roche, P. & Palmer, S. (2007). Noxious heat discrimination ability in persons with chronic pain: A psychophysical study. Physiotherapy, 93, Supplement 1, S136. Presented at the 15th International Congress of the World Confederation of Physical Therapists, Vancouver, Canada, June 02-06.